

ISOLATED SH3 GENES ASSOCIATED WITH MYELOPROLIFERATIVE
DISORDERS AND LEUKEMIA, AND USES THEREOF

RESEARCH SUPPORT

The research leading to the present invention was supported in part by the Clinical
5 Molecular Core grant NICHD P01HD17449 from the National Institutes of Health. The
government may have certain rights in the present invention.

FIELD OF THE INVENTION

10 The present invention relates to the isolated nucleic acids and corresponding amino acids
of a series of SH3 genes, analogs, fragments, mutants, and variants thereof. The
invention provides polypeptides, fusion proteins, chimerics, antisense molecules,
antibodies, and uses thereof. Also, this invention is directed to diagnostic methods of
determining whether a subject has a megakaryocytic abnormality, myeloproliferative
15 disorder, platelet disorder, hematopoietic disorder, or leukemia, or disorders associated
with abnormal neural development, and therapeutic treatments thereof.

BACKGROUND OF THE INVENTION

20 Down syndrome, caused by trisomy of human chromosome 21 (HSA21), is the most
common autosomal form of mental retardation. The first report describing an
association between Down syndrome (DS) and leukemia, which are an important cause
of morbidity and mortality worldwide, was presented in 1930. Since that time, the
increased incidence of acute leukemia in patients with DS has been clearly established.
25 However, the M7 subtype, AMKL, acute megakaryoblastic leukemia has been found
to be common in DS but relatively rare in non-DS. An instability in the control of
bone marrow proliferation has been hypothesized as a predisposing factor. The
incidence of acute myelogenous leukemia patients with DS has been noted by some to

be similar to that in children without mongolism. Chromosome 21 is a model for the study of human chromosomal aneuploidy, and the construction of its physical and transcriptional maps is a necessary step in understanding the molecular basis of aneuploidy-dependent phenotypes.

5

Human chromosome 21 has a nearly complete physical map with a well-characterized contiguous set of overlapping YACs spanning most of its length (Chumakov et al., 1992; Shimizu et al., 1995; Korenberg et al., 1995). The demand for sequence-ready contigs and clones for gene isolation efforts has prompted the construction of numerous
10 higher resolution contigs in cosmids (Patil et al., 1994; Soeda et al., 1995) and, more recently, in P1-derived artificial chromosomes (PACs; Oegawa et al. 1996 and Hubert et al. (1997) Genomics 41:218-226). Considerable mapping efforts exist in the region from CBR to D21S55 due to the common duplication of the region in partially trisomic individuals with several phenotypic features of DS, including mental retardation.
15 However, the distal and adjacent, 4- to 5-Mb D21S55 to MX1 region is also associated with DS-CHD as well as other characteristic features of DS (Korenberg et al., 1992, 1994).

Although full monosomy of chromosome 21 is usually lethal *in utero*, there are rare
20 cases of individuals with chromosome 21 deletions who survive. These individuals exhibit a characteristic subset of clinical features including psychomotor and growth retardation, congenital heart disease, holoprosencephaly, microphthalmia, skeletal malformations, and genital hypoplasia. Megakaryocytic abnormalities is added to this set and define a minimal "overlap" region for this feature through the clinical,
25 cytogenetic, and molecular analysis of four patients with overlapping deletions of chromosome 21 and thrombocytopenia.

Nonchimeric YACs span this interval with a few gaps but higher resolution physical maps are not available for most of the D21S55 to MX1 region. DEL21RW carries two
30 interstitial deletions, one in 21q21.3-22.1 defined by YAC 62G5 through YAC 760H5, and the second in 21q22.2, deleting IFNAR through CBR. DEL21LS carries an

- interstitial deletion of 21q22.1 from YAC 760H5 through the AML1 gene. Korenberg et al. reported that the deletion of patient DEL21HJ includes D21S93 through AML1. DEL21SV has a possible terminal deletion, 21q22.13-qter, extending from just proximal to D21S324 through D21S123. The common deleted region, or overlap
5 region, is therefore from D21S324 through AML1, a region of less than 2Mb that contains only three known genes, AML1, KCNE1, and UNO2. Bone marrow examination of two of the patients, DEL21HJ and Del 21RW, showed normocellular marrow with normal myelopoiesis, normal erythropoiesis, and small, dysplastic megakaryocytes with hypolobated nuclei. These two patients have decreased platelet
10 activation by agonists with normal platelet ultrastructures. All four patients have platelet dysfunction characterized by low platelet counts in the range of $31-113 \times 10^9$ /L. Further, all four subjects with chromosome 21 deletions that do not include this region have normal number of platelets.
- 15 A 3' fragment of SH3P17 gene was found in a study to isolate SH3 domain containing genes (Sparks et al. 1996, *Nature Biotechnology* 14:741). This was mapped to 21 or large sub-region of 21 by a number of groups by using database matches to the published sequence. Katsanis N, et al (Hum Genet 1997 Sep;100(3-4):477-480) utilized information generated by various EST sequencing projects to enrich the transcription
20 map of chromosome 21 and report the mapping of SH3P17 to 21q22.1 and the localisation of two genes previously mapped to HSA21 by Nagase and colleagues, KIAA0136 and KIAA0179 to 21q22.2 and 21q22.3 respectively. Chen H, and Antonarakis SE (Cytogenet Cell Genet 1997;78(3-4):213-215) identified portions of genes on human chromosome 21 and mapped the gene to YACs and cosmids within
25 21q22.1-->q22.2 between DNA markers D21S319 and D21S65 using hybridization and PCR amplification. Lastly, Guipponi et. al. 1998, *Genomics* 53:369-376 reported that they identified two isoforms of the human homolog of *Xenopus* Intersectin (ITSN) produced from alternate transcripts, the first of which, a short transcript is reportedly ubiquitously expressed, while the second longer transcript is exclusively expressed in
30 brain tissue. Later, Guipponi et. al. 1998 *Cytogenet Cell Genet.* 83:218-220 reported that they had identified the genomic structure, sequence and

precise mapping of the human intersectin gene and speculated that it may play a role in the determination of certain of the phenotypic characteristics of Down syndrome. The authors did not present evidence and corresponding observations or speculation regarding the role of the discovered genes apart from a possible relation to Down syndrome, and as such, are distinguishable from the research and discoveries embodied in the present invention.

The present invention provides the complete nucleotide sequence of several SH3 genes, including the SH3D1A gene and clones thereof, their association with platelet dysfunction and leukemia, including a part of the increased risk of leukemia seen in Down Syndrome, and with dysfunctions associated with neural development and particularly development in the CNS.

SUMMARY OF THE INVENTION

In one embodiment, this invention provides isolated nucleic acids which encode human SH3 genes such as SH3D1A and cDNA clones thereof, including also analogs, fragments, variants, and mutants, thereof. This invention is directed to an isolated nucleic acid encoding an amino acid sequence which forms one or more myristoylation sites in the EH domain and SH3 domain. This invention provides an isolated nucleic acid encoding an amino acid sequence which forms one or more EH domains and one or more SH3 domains. In one embodiment the nucleic acid which encodes an amino acid sequence which forms two EH domains and four SH3 domains. As shown in Figure 1 the nucleic acid encoding the amino acid sequence comprises one or more myristoylation sites in the EH domain and SH3 domain.

In one embodiment of this invention, the isolated nucleic acid encodes an amino acid sequence of the EH1 domain which is from amino acid sequence 15 to sequence 102. In another embodiment of this invention, the nucleic acid encodes an amino acid sequence of the EH2 domain which is from amino acid sequence 215 to sequence 310. In another embodiment of this invention, the nucleic acid encodes an amino acid sequence of the

SH3-1 domain which is from amino acid sequence 740 to sequence 800. In another embodiment of this invention, the nucleic acid encodes an amino acid sequence of the SH3-2 domain which is from amino acid sequence 908 to sequence 966. In another embodiment of this invention, the nucleic acid encodes an amino acid sequence of the SH3-3 domain which is from amino acid sequence 999 to sequence 1062. In another embodiment of this invention, the nucleic acid encodes an amino acid sequence of the SH3-4 domain which is from amino acid sequence 1080 to sequence 1138. In another embodiment of this invention, the nucleic acid encodes an amino acid sequence of the SH3-1 domain which is from amino acid sequence 740 to sequence 800. In a preferred embodiment, the nucleic acid encodes an amino acid sequence as set forth in SEQ. ID. NO. 2, and as set forth in Figures 5, 9, 11, 13 and 15.

This invention provides for an isolated nucleic acid which encodes SH3D1A, and clones thereof as set forth herein. The isolated nucleic acid may be DNA or RNA, specifically cDNA or genomic DNA. This isolated nucleic acid also encodes mutant SH3D1A or the wildtype protein. The isolated nucleic acid may also encode a human SH3D1A having substantially the same amino acid sequence as the sequence designated Figure 5. As used herein and in the claims, the terms nucleic acids encoding or expressing SH3D1A is intended to comprehend and include isolated nucleic acids that may have the sequence set forth in Figures 4, 8, 10, 12 or 14.

This invention is directed to a polypeptide comprising the amino acid sequence of a human SH3D1A or to a clone thereof. As used herein and in the claims, polypeptide or protein of SH3D1A is intended to comprehend and include polypeptides that comprise or otherwise correspond to those set forth in Figures 9, 11, 13, or 15 herein, or analogs or fragments thereof. Further, polyclonal and monoclonal antibodies which specifically bind to the polypeptide are disclosed and chimeric (bi-specific) antibodies are likewise contemplated.

This invention provides a method for determining whether a subject carries a mutation in the SH3D1A gene which comprises: (a) obtaining an appropriate nucleic acid sample

from the subject; and (b) determining whether the nucleic acid sample from step (a) is, or is derived from, a nucleic acid which encodes mutant SH3D1A so as to thereby determine whether a subject carries a mutation in the SH3D1A gene.

- 5 This invention provides a method for determining whether a subject has a megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, or leukemia, or a neural disorder which comprises: (a) obtaining an appropriate sample from the subject; and (b) contacting the sample with the antibody so as to thereby determine whether a subject has the megakaryocytic abnormality, myeloproliferative disorder,
10 platelet disorder, leukemia or neural disorder.

- This invention provides a method for determining whether a subject has a predisposition for a megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, leukemia, or a neural disorder, which comprises: (a) obtaining an appropriate nucleic
15 acid sample from the subject; and (b) determining whether the nucleic acid sample from step (a) is, or is derived from, a nucleic acid which encodes SH3D1A so as to thereby determine whether a subject has a predisposition for a megakaryocytic abnormality, myeloproliferative disorder, platelet disorder or leukemia, or a neural disorder.

- 20 This invention provides a method for determining whether a subject has a megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, leukemia, or a neural disorder, which comprises: (a) obtaining an appropriate nucleic acid sample from the subject; and (b) determining whether the nucleic acid sample from step (a) is, or is derived from, a nucleic acid which encodes the human SH3D1A so as to thereby
25 determine whether a subject has megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, leukemia, or a neural disorder,.

- This invention provides a method for screening a tumor sample from a human subject for a somatic alteration in a SH3D1A gene in said tumor which comprises gene comparing
30 a first sequence selected from the group consisting of a SH3D1A gene from said tumor sample, SH3D1A RNA from said tumor sample and SH3D1A cDNA made from mRNA

from said tumor sample with a second sequence selected from the group consisting of SH3D1A gene from a nontumor sample of said subject, SH3D1A RNA from said nontumor sample and SH3D1A cDNA made from mRNA from said nontumor sample, wherein a difference in the sequence of the SH3D1A gene, SH3D1A RNA or SH3D1A
5 cDNA from said tumor sample from the sequence of the SH3D1A gene, SH3D1A RNA or SH3D1A cDNA from said nontumor sample indicates a somatic alteration in the SH3D1A gene in said tumor sample.

This invention provides a method for monitoring the progress and adequacy of treatment
10 in a subject who has received treatment for a megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, leukemia or an abnormal neural condition which comprises monitoring the level of nucleic acid encoding the human SH3D1A at various stages of treatment.

15 The present invention provides the means necessary for production of gene-based therapies directed at cancer cells; diagnosis of the predisposition to, and diagnosis and treatment of megakaryocytic abnormality, hematopoietic disorders, myeloproliferative disorder, platelet disorder, Down Syndrome, leukemia, other disorders based in whole or in part from neural abnormalities or dysfunctions; and prenatal diagnosis and
20 treatment of tumors. These therapeutic agents may take the form of polynucleotides comprising all or a portion of the SH3D1A gene placed in appropriate vectors or delivered to target cells in more direct ways such that the function of the SH3D1A protein is reconstituted. Therapeutic agents may also take the form of polypeptides based on either a portion of, or the entire protein sequence of SH3D1A.

25

This invention provides a pharmaceutical composition comprising an amount of the polypeptide of the human SH3D1A as defined herein, and a pharmaceutically effective carrier or diluent.

30 This invention provides a method of treating a subject having megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, leukemia or neural

abnormality or dysfunction, which comprises introducing the isolated nucleic acid into the subject under conditions such that the nucleic acid expresses SH3D1A, so as to thereby treat the subject.

- 5 This invention provides a method of treating a subject having megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, leukemia, or neural abnormality or dysfunction, which comprises administration to the subject a therapeutically effective amount of the pharmaceutical composition to the subject.
- 10 Lastly, the present invention also provides kits for detecting in an analyte at least one oligonucleotide comprising the SH3D1A gene, or a portion thereof, the kits comprising polynucleotide complementary to the SH3D1A gene, a fragment, binding partner, analog or other portion thereof, gene packaged in a suitable container, and instructions for its use.

15

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1. Human SH3D1A structure and homology

20 **FIGURE 2.** SH3D1A domain structure and homologies - human vs. Xenopus

FIGURE 3. Region of chromosome 21 responsible for megakaryocytic abnormalities.

FIGURE 4. Nucleic acid sequence of human SH3D1A.

25

FIGURE 5. Amino acid sequence of human SH3D1A.

FIGURE 6. Northern Blot of SH3D1A expressed in heart, brain, placenta, lung, liver, muscle, kidney and pancreas.

30

FIGURE 7. Map presenting four cDNA clones in accordance with the invention,

including length and protein domains.

FIGURE 8. Nucleic acid sequence of cDNA clone also identified herein as Clone #21.

5 **FIGURE 9.** Amino acid sequence of Clone #21. Upper part of Figure presents translated protein sequence; lower portion of Figure presents whole protein sequence.

FIGURE 10. Nucleic acid sequence of cDNA clone also identified herein as Clone #11.

10

FIGURE 11. Amino acid sequence of Clone #11. Upper part of Figure presents translated protein sequence; lower portion of Figure presents whole protein sequence.

15

FIGURE 12. Nucleic acid sequence of cDNA clone also identified herein as Clone #5.

FIGURE 13. Amino acid sequence of Clone #5. Upper part of Figure presents translated protein sequence; lower portion of Figure presents whole protein sequence.

20

FIGURE 14. Nucleic acid sequence of cDNA clone also identified herein as Clone #9.

FIGURE 15. Amino acid sequence of Clone #5. Upper part of Figure presents translated protein sequence; lower portion of Figure presents whole protein sequence.

25

FIGURE 16. Tissue immunochemical staining on mouse embryo (Day 9) showing ITSN expression in neural blasts during migration and formation in CNS.

30

FIGURE 17. Summary of Studies on ITSN:

I. Gene sequence: First line showing the scale of ITSN cDNA; Second line showing the total numbers of the exons and the positions of each exon located.

5 **II. Protein domains vs nucleotide sequence:** ITSN was predicted consists of 11 protein domains as listed on the map - 2 EH domains, 5 SH3 domains and 1 of each GEF, pH and C2 domains. Their relative positions on the cDNA level were numbered under each domain.

10 **III. Gene expression of human adult and fetal tissues:** This part summarized the Northern blot results showing ITSN was ubiquitously expressed with extensive alternative splicing generating tissue and developmental stage-specific expression.

FIGURE 18. Sequence comparisons between nucleic acid molecules of present invention, and Intersectins (ITSN), including a consensus sequence.

15

DETAILED DESCRIPTION OF THE INVENTION

The present invention discloses a family of SH3 genes, and particularly, a novel SH3D1A gene, and clones, and corresponding proteins, both translated and full length, which
20 SH3D1A gene is on chromosome 21, and that contributes to the development of platelets and the pathogenesis of leukemias, both in general and in particular those involving the megakaryocytic lineage. The invention provides methods useful for diagnosing and treating the following: acute leukemias, thrombocytopenia, megakaryocytic abnormality, hematopoietic disorders, myeloproliferative disorder, platelet disorder, leukemia,
25 leukemia in Down syndrome, leukemia, platelet disorder on chromosome 21, low platelets in deletion for 21, association of gains in chromosome 21 with leukemias and disorders associated with associated with megakaryocytic dysfunction; and neural abnormalities, dysfunctions and disorders, including brain malformations and corresponding cognitive dysfunctions, microcephaly, lissencephaly, colpocephaly,
30 holoprosencephaly.

This invention provides an isolated nucleic acid which encodes a human SH3D1A, as defined hereinabove, including analogs, such as the nucleic acids set forth in Figures 8, 10, 12 and 14, fragments, presented herein by way of non-limiting example, variants, and mutants, thereof. In one embodiment the nucleic acid has a nucleotide sequence having
5 at least 85% similarity with the nucleic acid coding sequence of SEQ ID NO: 1. This invention is directed to an isolated nucleic acid encoding an amino acid sequence which forms one or more myristoylation sites in the EH domain and SH3 domain. This invention provides a isolated nucleic acid encoding an amino acid sequence which forms one or more EH domains and one or more SH3 domains. In one embodiment the nucleic
10 acid which encodes an amino acid sequence which forms two EH domains and four SH3 domains. As show in Figure 1 the nucleic acid encoding the amino acid sequence comprising one or more myristoylation sites in the EH domain and SH3 domain.

In one embodiment of this invention, the isolated nucleic acid encodes an amino acid
15 sequence of the EH1 domain which corresponds to the following regions: amino acid sequence 15 to sequence 102. In another embodiment of this invention, the nucleic acid encodes an amino acid sequence of the EH2 domain which is from amino acid sequence 215 to sequence 310. In another embodiment of this invention, the nucleic acid encodes an amino acid sequence of the SH3-1 domain which is from amino acid sequence 740 to
20 sequence 800. In another embodiment of this invention, the nucleic acid encodes an amino acid sequence of the SH3-2 domain which is from amino acid sequence 908 to sequence 966. In another embodiment of this invention, the nucleic acid encodes an amino acid sequence of the SH3-3 domain which is from amino acid sequence 999 to sequence 1062. In another embodiment of this invention, the nucleic acid encodes an
25 amino acid sequence of the SH3-4 domain which is from amino acid sequence 1080 to sequence 1138. In another embodiment of this invention, the nucleic acid encodes an amino acid sequence of the SH3-1 domain which is from amino acid sequence 740 to sequence 800. In a preferred embodiment, the nucleic acid encodes an amino acid sequence as set forth in Figure 5, or the corresponding analogs set forth in Figures 9, 11,
30 13 and 15, presented herein by way of non-limiting example. This invention contemplates nucleic acid or amino acid sequences which correspond to the SH3D1A

gene, analogs, fragments, variants, mutants thereof. The corresponding nucleic acids or amino acids may be based on nucleic acid, or amino acid sequence as disclosed herein; or based on the structurally or functionally of the EH and SH3 domains which define the SH3D1A gene.

5

This invention provides for an isolated nucleic acid which encodes SH3D1A. This isolated nucleic acid may be DNA or RNA, specifically cDNA or genomic DNA. This isolated nucleic acid also encodes mutant SH3D1A or the wildtype protein. The isolated nucleic acid may also encode a human SH3D1A having substantially the same amino acid sequence as the sequence designated Figure 5. Specifically the isolated nucleic acid has the sequence designated Figure 4.

10

This invention provides for a replicable vector comprising the isolated nucleic acid molecule of the DNA virus. The vector includes, but is not limited to: a plasmid, cosmid, λ phage or yeast artificial chromosome (YAC) which contains at least a portion of the isolated nucleic acid molecule. As an example to obtain these vectors, insert and vector DNA can both be exposed to a restriction enzyme to create complementary ends on both molecules which base pair with each other and are then ligated together with DNA ligase. Alternatively, linkers can be ligated to the insert DNA which correspond to a restriction site in the vector DNA, which is then digested with the restriction enzyme which cuts at that site. Other means are also available and known to an ordinary skilled practitioner.

15

20

Regulatory elements required for expression include promoter or enhancer sequences to bind RNA polymerase and transcription initiation sequences for ribosome binding. For example, a bacterial expression vector includes a promoter such as the lac promoter and for transcription initiation the Shine-Dalgarno sequence and the start codon AUG. Similarly, a eukaryotic expression vector includes a heterologous or homologous promoter for RNA polymerase II, a downstream polyadenylation signal, the start codon AUG, and a termination codon for detachment of the ribosome. Such vectors may be obtained commercially or assembled from the sequences described by methods well-known in the art, for example the methods described above for constructing vectors in

25

30

general.

This invention provides a host cell containing the above vector. The host cell may contain the isolated DNA molecule artificially introduced into the host cell. The host cell
5 may be a eukaryotic or bacterial cell (such as E.coli), yeast cells, fungal cells, insect cells and animal cells. Suitable animal cells include, but are not limited to Vero cells, HeLa cells, Cos cells, CV1 cells and various primary mammalian cells.

The term "vector", refers to viral expression systems, autonomous self-replicating
10 circular DNA (plasmids), and includes both expression and nonexpression plasmids. Where a recombinant microorganism or cell culture is described as hosting an "expression vector," this includes both extrachromosomal circular DNA and DNA that has been incorporated into the host chromosome(s). Where a vector is being maintained by a host cell, the vector may either be stably replicated by the cells during mitosis as an
15 autonomous structure, or is incorporated within the host's genome.

The term "plasmid" refers to an autonomous circular DNA molecule capable of replication in a cell, and includes both the expression and nonexpression types. Where a recombinant microorganism or cell culture is described as hosting an "expression
20 plasmid", this includes latent viral DNA integrated into the host chromosome(s). Where a plasmid is being maintained by a host cell, the plasmid is either being stably replicated by the cells during mitosis as an autonomous structure or is incorporated within the host's genome.

25 The following terms are used to describe the sequence relationships between two or more nucleic acid molecules or polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of
30 a full-length cDNA or gene sequence given in a sequence listing or may comprise a complete cDNA or gene sequence.

Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and Waterman (1981) *Adv. Appl. Math.* 2:482, by the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity method of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. (USA)* 85:2444, or by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI).

"Substantial identity" or "substantial sequence identity" mean that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap which share at least 90 percent sequence identity, preferably at least 95 percent sequence identity, more preferably at least 99 percent sequence identity or more. "Percentage amino acid identity" or "percentage amino acid sequence identity" refers to a comparison of the amino acids of two polypeptides which, when optimally aligned, have approximately the designated percentage of the same amino acids. For example, "95% amino acid identity" refers to a comparison of the amino acids of two polypeptides which when optimally aligned have 95% amino acid identity. Preferably, residue positions which are not identical differ by conservative amino acid substitutions. For example, the substitution of amino acids having similar chemical properties such as charge or polarity are not likely to effect the properties of a protein. Examples include glutamine for asparagine or glutamic acid for aspartic acid.

The phrase "nucleic acid molecule encoding" refers to a nucleic acid molecule which directs the expression of a specific protein or peptide. The nucleic acid sequences include both the DNA strand sequence that is transcribed into RNA and the RNA sequence that is translated into protein. The nucleic acid molecule include both the full length nucleic acid sequences as well as non-full length sequences derived from the full length protein. It being further understood that the sequence includes the degenerate codons of the native sequence or sequences which may be introduced to provide codon preference in a specific host cell.

This invention provides a nucleic acid having a sequence complementary to the sequence of the isolated nucleic acid of the human SH3D1A gene. Specifically, this invention provides an oligonucleotide of at least 15 nucleotides capable of specifically hybridizing with a sequence of nucleotides present within a nucleic acid which encodes the human
5 SH3D1A. In one embodiment the nucleic acid is DNA or RNA. In another embodiment the oligonucleotide is labeled with a detectable marker. In another embodiment the oligonucleotide is a radioactive isotope, a fluorophor or an enzyme.

Oligonucleotides which are complementary may be obtained as follows: The polymerase
10 chain reaction is then carried out using the two primers. See *PCR Protocols: A Guide to Methods and Applications* [74]. Following PCR amplification, the PCR-amplified regions of a viral DNA can be tested for their ability to hybridize to the three specific nucleic acid probes listed above. Alternatively, hybridization of a viral DNA to the above nucleic acid probes can be performed by a Southern blot procedure without viral
15 DNA amplification and under stringent hybridization conditions as described herein.

Oligonucleotides for use as probes or PCR primers are chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage and Carruthers [19] using an automated synthesizer, as described in Needham-VanDevanter
20 [69]. Purification of oligonucleotides is by either native acrylamide gel electrophoresis or by anion-exchange HPLC as described in Pearson, J.D. and Regnier, F.E. [75A]. The sequence of the synthetic oligonucleotide can be verified using the chemical degradation method of Maxam, A.M. and Gilbert, W. [63].

25 High stringency hybridization conditions are selected at about 5° C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is at least about 0.02 molar at pH 7 and the
30 temperature is at least about 60°C. As other factors may significantly affect the stringency of hybridization, including, among others, base composition and size of the

complementary strands, the presence of organic solvents, ie. salt or formamide concentration, and the extent of base mismatching, the combination of parameters is more important than the absolute measure of any one. For Example high stringency may be attained for example by overnight hybridization at about 68°C in a 6x SSC solution, washing at room temperature with 6x SSC solution, followed by washing at about 68°C in a 6x SSC in a 0.6x SSX solution.

Hybridization with moderate stringency may be attained for example by: 1) filter pre-hybridizing and hybridizing with a solution of 3x sodium chloride, sodium citrate (SSC), 50% formamide, 0.1M Tris buffer at Ph 7.5, 5x Denhardt's solution; 2.) pre-hybridization at 37°C for 4 hours; 3) hybridization at 37°C with amount of labelled probe equal to 3,000,000 cpm total for 16 hours; 4) wash in 2x SSC and 0.1% SDS solution; 5) wash 4x for 1 minute each at room temperature at 4x at 60°C for 30 minutes each; and 6) dry and expose to film.

The phrase "selectively hybridizing to" refers to a nucleic acid probe that hybridizes, duplexes or binds only to a particular target DNA or RNA sequence when the target sequences are present in a preparation of total cellular DNA or RNA. By selectively hybridizing it is meant that a probe binds to a given target in a manner that is detectable in a different manner from non-target sequence under high stringency conditions of hybridization. In a different "Complementary" or "target" nucleic acid sequences refer to those nucleic acid sequences which selectively hybridize to a nucleic acid probe. Proper annealing conditions depend, for example, upon a probe's length, base composition, and the number of mismatches and their position on the probe, and must often be determined empirically. For discussions of nucleic acid probe design and annealing conditions, see, for example, Sambrook *et al.*, [81] or Ausubel, F., *et al.*, [8].

It will be readily understood by those skilled in the art and it is intended here, that when reference is made to particular sequence listings, such reference includes sequences which substantially correspond to its complementary sequence and those described including allowances for minor sequencing errors, single base changes, deletions,

substitutions and the like, including the clonal variants set forth herein, such that any such sequence variation corresponds to the nucleic acid sequence of the pathogenic organism or disease marker to which the relevant sequence listing relates.

- 5 Nucleic acid probe technology is well known to those skilled in the art who readily appreciate that such probes may vary greatly in length and may be labeled with a detectable label, such as a radioisotope or fluorescent dye, to facilitate detection of the probe. DNA probe molecules may be produced by insertion of a DNA molecule having the full-length or a fragment of the isolated nucleic acid molecule of the DNA virus into
- 10 suitable vectors, such as plasmids or bacteriophages, followed by transforming into suitable bacterial host cells, replication in the transformed bacterial host cells and harvesting of the DNA probes, using methods well known in the art. Alternatively, probes may be generated chemically from DNA synthesizers.
- 15 RNA probes may be generated by inserting the full length or a fragment of the isolated nucleic acid molecule of the DNA virus downstream of a bacteriophage promoter such as T3, T7 or SP6. Large amounts of RNA probe may be produced by incubating the labeled nucleotides with a linearized isolated nucleic acid molecule of the DNA virus or its fragment where it contains an upstream promoter in the presence of the appropriate
- 20 RNA polymerase.

- As defined herein nucleic acid probes may be DNA or RNA fragments. DNA fragments can be prepared, for example, by digesting plasmid DNA, or by use of PCR, or synthesized by either the phosphoramidite method described by Beaucage and
- 25 Carruthers, [19], or by the triester method according to Matteucci, *et al.*, [62], both incorporated herein by reference. A double stranded fragment may then be obtained, if desired, by annealing the chemically synthesized single strands together under appropriate conditions or by synthesizing the complementary strand using DNA polymerase with an appropriate primer sequence. Where a specific sequence for a
- 30 nucleic acid probe is given, it is understood that the complementary strand is also identified and included. The complementary strand will work equally well in situations

where the target is a double-stranded nucleic acid. It is also understood that when a specific sequence is identified for use a nucleic probe, a subsequence of the listed sequence which is 25 basepairs or more in length is also encompassed for use as a probe.

- 5 The DNA molecules of the subject invention also include DNA molecules coding for polypeptide analogs, fragments or derivatives of antigenic polypeptides which differ from naturally-occurring forms in terms of the identity or location of one or more amino acid residues (deletion analogs containing less than all of the residues specified for the protein, substitution analogs wherein one or more residues specified are replaced by other
- 10 residues and addition analogs where in one or more amino acid residues is added to a terminal or medial portion of the polypeptides) and which share some or all properties of naturally-occurring forms. These molecules include: the incorporation of codons "preferred" for expression by selected non-mammalian hosts; the provision of sites for cleavage by restriction endonuclease enzymes; and the provision of additional initial,
- 15 terminal or intermediate DNA sequences that facilitate construction of readily expressed vectors.

- Also, this invention provides an antisense molecule capable of specifically hybridizing with the isolated nucleic acid of the human SH3D1A gene. This invention provides an
- 20 antagonist capable of blocking the expression of the peptide or polypeptide encoded by the isolated DNA molecule. In one embodiment the antagonist is capable of hybridizing with a double stranded DNA molecule. In another embodiment the antagonist is a triplex oligonucleotide capable of hybridizing to the DNA molecule. In another embodiment the triplex oligonucleotide is capable of binding to at least a portion of the isolated DNA
- 25 molecule with a nucleotide sequence..

- The antisense molecule may be DNA or RNA or variants thereof (i.e. DNA or RNA with a protein backbone). The present invention extends to the preparation of antisense nucleotides and ribozymes that may be used to interfere with the expression of the
- 30 receptor recognition proteins at the translation of a specific mRNA, either by masking that MRNA with an antisense nucleic acid or cleaving it with a ribozyme.

Antisense nucleic acids are DNA or RNA molecules that are complementary to at least a portion of a specific mRNA molecule. In the cell, they hybridize to that mRNA, forming a double stranded molecule. The cell does not translate an mRNA in this double-stranded form. Therefore, antisense nucleic acids interfere with the expression of mRNA into protein.

Antisense nucleotides or polynucleotide sequences are useful in preventing or diminishing the expression of the SH3D1A gene, as will be appreciated by those skilled in the art. For example, polynucleotide vectors containing all or a portion of the SH3D1A gene or other sequences from the SH3D1A region (particularly those flanking the SH3D1A gene) may be placed under the control of a promoter in an antisense orientation and introduced into a cell. Expression of such an antisense construct within a cell will interfere with SH3D1A transcription and/or translation and/or replication. Oligomers of about fifteen nucleotides and molecules that hybridize to the AUG initiation codon are particularly efficient, since they are easy to synthesize and are likely to pose fewer problems than larger molecules upon introduction to cells.

This invention provides a transgenic nonhuman mammal which comprises at least a portion of the isolated DNA molecule introduced into the mammal at an embryonic stage. Methods of producing a transgenic nonhuman mammal are known to those skilled in the art.

This invention also provides a method of producing a polypeptide encoded by isolated DNA molecule, which comprises growing the above host vector system under suitable conditions permitting production of the polypeptide and recovering the polypeptide so produced.

This invention provides a polypeptide comprising the amino acid sequence of a human SH3D1A. In one embodiment, the amino acid sequence is set forth in Figure 5. Further, the isolated polypeptide encoded by the isolated DNA molecule may be linked to a second polypeptide encoded by a nucleic acid molecule to form a fusion protein by

expression in a suitable host cell. In one embodiment the second nucleic acid molecule encodes beta-galactosidase. Other nucleic acid molecules which are used to form a fusion protein are known to those skilled in the art.

- 5 This invention provides an antibody which specifically binds to the polypeptide encoded by the isolated DNA molecule. In one embodiment the antibody is a monoclonal antibody. In another embodiment the antibody is a polyclonal antibody. The antibody or DNA molecule may be labelled with a detectable marker including, but not limited to: a radioactive label, or a colorimetric, a luminescent, or a fluorescent marker, or gold.
- 10 Radioactive labels include, but are not limited to: ^3H , ^{14}C , ^{32}P , ^{33}P , ^{35}S , ^{36}Cl , ^{51}Cr , ^{57}Co , ^{59}Co , ^{59}Fe , ^{90}Y , ^{125}I , ^{131}I , and ^{186}Re . Fluorescent markers include but are not limited to: fluorescein, rhodamine and auramine. Colorimetric markers include, but are not limited to: biotin, and digoxigenin. Methods of producing the polyclonal or monoclonal antibody are known to those of ordinary skill in the art.
- 15 Further, the antibody or nucleic acid molecule complex may be detected by a second antibody which may be linked to an enzyme, such as alkaline phosphatase or horseradish peroxidase. Other enzymes which may be employed are well known to one of ordinary skill in the art.
- 20 "Specifically binds to an antibody" or "specifically immunoreactive with", when referring to a protein or peptide, refers to a binding reaction which is determinative of the presence of the SH3D1A of the invention in the presence of a heterogeneous population of proteins and other biologics including viruses other than the SH3D1A. Thus, under
- 25 designated immunoassay conditions, the specified antibodies bind to the SH3D1A antigens and do not bind in a significant amount to other antigens present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, antibodies raised to the human SH3D1A immunogen described herein can be selected to obtain antibodies
- 30 specifically immunoreactive with the SH3D1A proteins and not with other proteins. These antibodies recognize proteins homologous to the human SH3D1A protein. A

variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane [32] for a description of
5 immunoassay formats and conditions that can be used to determine specific immunoreactivity.

This invention provides a method to select specific regions on the polypeptide encoded by the isolated DNA molecule of the DNA virus to generate antibodies. The protein
10 sequence may be determined from the cDNA sequence. Amino acid sequences may be analyzed by methods well known to those skilled in the art to determine whether they produce hydrophobic or hydrophilic regions in the proteins which they build. In the case of cell membrane proteins, hydrophobic regions are well known to form the part of the protein that is inserted into the lipid bilayer of the cell membrane, while hydrophilic
15 regions are located on the cell surface, in an aqueous environment. Usually, the hydrophilic regions will be more immunogenic than the hydrophobic regions. Therefore the hydrophilic amino acid sequences may be selected and used to generate antibodies specific to polypeptide encoded by the isolated nucleic acid molecule encoding the DNA virus. The selected peptides may be prepared using commercially available machines.
20 As an alternative, DNA, such as a cDNA or a fragment thereof, may be cloned and expressed and the resulting polypeptide recovered and used as an immunogen.

Polyclonal antibodies against these peptides may be produced by immunizing animals using the selected peptides. Monoclonal antibodies are prepared using hybridoma
25 technology by fusing antibody producing B cells from immunized animals with myeloma cells and selecting the resulting hybridoma cell line producing the desired antibody. Alternatively, monoclonal antibodies may be produced by *in vitro* techniques known to a person of ordinary skill in the art. Also as set forth earlier herein, chimeric (bi-specific) antibodies may be prepared by techniques well known in the art, and are likewise
30 contemplated herein. Any and all of these antibodies are useful to detect the expression of polypeptide encoded by the isolated DNA molecule of the DNA virus in living

animals, in humans, or in biological tissues or fluids isolated from animals or humans.

The antibodies may be detectably labeled, utilizing conventional labeling techniques well-known to the art. Thus, the antibodies may be radiolabeled using, for example, 5 radioactive isotopes such as ^3H , ^{125}I , ^{131}I , and ^{35}S . The antibodies may also be labeled using fluorescent labels, enzyme labels, free radical labels, or bacteriophage labels, using techniques known in the art. Typical fluorescent labels include fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, and Texas Red.

10 Since specific enzymes may be coupled to other molecules by covalent links, the possibility also exists that they might be used as labels for the production of tracer materials. Suitable enzymes include alkaline phosphatase, beta-galactosidase, glucose-6-phosphate dehydrogenase, maleate dehydrogenase, and peroxidase. Two principal types of enzyme immunoassay are the enzyme-linked immunosorbent assay 15 (ELISA), and the homogeneous enzyme immunoassay, also known as enzyme-multiplied immunoassay (EMIT, Syva Corporation, Palo Alto, CA). In the ELISA system, separation may be achieved, for example, by the use of antibodies coupled to a solid phase. The EMIT system depends on deactivation of the enzyme in the tracer-antibody complex; the activity can thus be measured without the need for a separation step.

20

Additionally, chemiluminescent compounds may be used as labels. Typical chemiluminescent compounds include luminol, isoluminol, aromatic acridinium esters, imidazoles, acridinium salts, and oxalate esters. Similarly, bioluminescent compounds may be utilized for labelling, the bioluminescent compounds including luciferin, 25 luciferase, aequorin, and fluorescent proteins such as green fluorescent protein (GFP). Once labeled, the antibody may be employed to identify and quantify immunologic counterparts (antibody or antigenic polypeptide) utilizing techniques well-known to the art.

30 A description of a radioimmunoassay (RIA) may be found in *Laboratory Techniques in Biochemistry and Molecular Biology* [52], with particular reference to the chapter entitled

"An Introduction to Radioimmune Assay and Related Techniques" by Chard, T., incorporated by reference herein. A description of general immunometric assays of various types can be found in the following U.S. Pat. Nos. 4,376,110 (David *et al.*) or 4,098,876 (Piasio).

5

One can use immunoassays to detect for the SH3D1A gene, specific peptides, or for antibodies to the virus or peptides. A general overview of the applicable technology is in Harlow and Lane [32], incorporated by reference herein.

- 10 In one embodiment, antibodies to the human SH3D1A can be used to detect the agent in the sample. In brief, to produce antibodies to the agent or peptides, the sequence being targeted is expressed in transfected cells, preferably bacterial cells, and purified. The product is injected into a mammal capable of producing antibodies. Either monoclonal or polyclonal antibodies (as well as any recombinant antibodies) specific for the gene
- 15 product can be used in various immunoassays. Such assays include competitive immunoassays, radioimmunoassays, Western blots, ELISA, indirect immunofluorescent assays and the like. For competitive immunoassays, see Harlow and Lane [32] at pages 567-573 and 584-589.
- 20 In a further embodiment of this invention, commercial test kits suitable for use by a medical specialist may be prepared to determine the presence or absence of predetermined binding activity or predetermined binding activity capability to suspected target cells. In accordance with the testing techniques discussed above, one class of such kits will contain at least the labeled polypeptide or its binding partner, for instance an
- 25 antibody specific thereto, and directions, of course, depending upon the method selected, *e.g.*, "competitive," "sandwich," "DASP" and the like. The kits may also contain peripheral reagents such as buffers, stabilizers, etc.

- Monoclonal antibodies or recombinant antibodies may be obtained by various techniques
- 30 familiar to those skilled in the art. Briefly, spleen cells or other lymphocytes from an animal immunized with a desired antigen are immortalized, commonly by fusion with

a myeloma cell (see, Kohler and Milstein [50], incorporated herein by reference). Alternative methods of immortalization include transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods well known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. New techniques using recombinant phage antibody expression systems can also be used to generate monoclonal antibodies. See for example: McCafferty, J *et al.* [64]; Hoogenboom, H.R. *et al.* [39]; and Marks, J.D. *et al.* [60].

Such peptides may be produced by expressing the specific sequence in a recombinantly engineered cell such as bacteria, yeast, filamentous fungal, insect (especially employing baculoviral vectors), and mammalian cells. Those of skill in the art are knowledgeable in the numerous expression systems available for expression of herpes virus protein.

Briefly, the expression of natural or synthetic nucleic acids encoding viral protein will typically be achieved by operably linking the desired sequence or portion thereof to a promoter (which is either constitutive or inducible), and incorporated into an expression vector. The vectors are suitable for replication or integration in either prokaryotes or eukaryotes. Typical cloning vectors contain antibiotic resistance markers, genes for selection of transformants, inducible or regulatable promoter regions, and translation terminators that are useful for the expression of viral genes.

Methods for the expression of cloned genes in bacteria are also well known. In general, to obtain high level expression of a cloned gene in a prokaryotic system, it is advisable to construct expression vectors containing a strong promoter to direct mRNA transcription. The inclusion of selection markers in DNA vectors transformed in *E. coli* is also useful. Examples of such markers include genes specifying resistance to antibiotics. See [81] *supra*, for details concerning selection markers and promoters for use in *E. coli*. Suitable eukaryote hosts may include plant cells, insect cells, mammalian

cells, yeast, and filamentous fungi.

- The peptides derived from the nucleic acids, peptide fragments are produced by recombinant technology may be purified by standard techniques well known to those of skill in the art. Recombinantly produced sequences can be directly expressed or expressed as a fusion protein. The protein is then purified by a combination of cell lysis (e.g., sonication) and affinity chromatography. For fusion products, subsequent digestion of the fusion protein with an appropriate proteolytic enzyme releases the desired peptide.
- 10 The proteins may be purified to substantial purity by standard techniques well known in the art, including selective precipitation with such substances as ammonium sulfate, column chromatography, immunopurification methods, and others. See, for instance, Scopes, R. [84], incorporated herein by reference.
- 15 This invention is directed to analogs of the isolated nucleic acid and polypeptide which comprise the amino acid sequence as set forth above. The analog may have an N-terminal methionine or an N-terminal polyhistidine optionally attached to the N or COOH terminus of the polypeptide which comprise the amino acid sequence.
- 20 In another embodiment, this invention contemplates peptide fragments of the polypeptide which result from proteolytic digestion products of the polypeptide. In another embodiment, the derivative of the polypeptide has one or more chemical moieties attached thereto. In another embodiment the chemical moiety is a water soluble polymer. In another embodiment the chemical moiety is polyethylene glycol. In another
- 25 embodiment the chemical moiety is mon-, di-, tri- or tetrapegylated. In another embodiment the chemical moiety is N-terminal monopegylated.

Attachment of polyethylene glycol (PEG) to compounds is particularly useful because PEG has very low toxicity in mammals (Carpenter et al., 1971). For example, a PEG adduct of adenosine deaminase was approved in the United States for use in humans for

30 the treatment of severe combined immunodeficiency syndrome. A second advantage

afforded by the conjugation of PEG is that of effectively reducing the immunogenicity and antigenicity of heterologous compounds. For example, a PEG adduct of a human protein might be useful for the treatment of disease in other mammalian species without the risk of triggering a severe immune response. The compound of the present invention
5 may be delivered in a microencapsulation device so as to reduce or prevent an host immune response against the compound or against cells which may produce the compound. The compound of the present invention may also be delivered microencapsulated in a membrane, such as a liposome.

10 Numerous activated forms of PEG suitable for direct reaction with proteins have been described. Useful PEG reagents for reaction with protein amino groups include active esters of carboxylic acid or carbonate derivatives, particularly those in which the leaving groups are N-hydroxysuccinimide, p-nitrophenol, imidazole or 1-hydroxy-2-nitrobenzene-4-sulfonate. PEG derivatives containing maleimido or haloacetyl groups
15 are useful reagents for the modification of protein free sulfhydryl groups. Likewise, PEG reagents containing amino hydrazine or hydrazide groups are useful for reaction with aldehydes generated by periodate oxidation of carbohydrate groups in proteins.

In one embodiment, the amino acid residues of the polypeptide described herein are
20 preferred to be in the "L" isomeric form. In another embodiment, the residues in the "D" isomeric form can be substituted for any L-amino acid residue, as long as the desired functional property of lectin activity is retained by the polypeptide. NH₂ refers to the free amino group present at the amino terminus of a polypeptide. COOH refers to the free carboxy group present at the carboxy terminus of a polypeptide. Abbreviations used
25 herein are in keeping with standard polypeptide nomenclature, *J. Biol. Chem.*, **243**:3552-59 (1969).

It should be noted that all amino-acid residue sequences are represented herein by formulae whose left and right orientation is in the conventional direction of amino-
30 terminus to carboxy-terminus. Furthermore, it should be noted that a dash at the beginning or end of an amino-acid residue sequence indicates a peptide bond to a further

sequence of one or more amino acid residues.

Synthetic polypeptides, prepared using the well known techniques of solid phase, liquid phase, or peptide condensation techniques, or any combination thereof, can include
5 natural and unnatural amino acids. Amino acids used for peptide synthesis may be standard Boc (N^α -amino protected N^α -t-butyloxycarbonyl) amino acid resin with the standard deprotecting, neutralization, coupling and wash protocols of the original solid phase procedure of Merrifield (1963, J. Am. Chem. Soc. 85:2149-2154), or the base-labile N^α -amino protected 9-fluorenylmethoxycarbonyl (Fmoc) amino acids first
10 described by Carpino and Han (1972, J. Org. Chem. 37:3403-3409). Thus, polypeptide of the invention may comprise D-amino acids, a combination of D- and L-amino acids, and various "designer" amino acids (e.g., β -methyl amino acids, $C\alpha$ -methyl amino acids, and $N\alpha$ -methyl amino acids, etc.) to convey special properties. Synthetic amino acids include ornithine for lysine, fluorophenylalanine for phenylalanine, and norleucine for
15 leucine or isoleucine. Additionally, by assigning specific amino acids at specific coupling steps, α -helices, β turns, β sheets, γ -turns, and cyclic peptides can be generated.

In one aspect of the invention, the peptides may comprise a special amino acid at the C-
20 terminus which incorporates either a CO_2H or $CONH_2$ side chain to simulate a free glycine or a glycine-amide group. Another way to consider this special residue would be as a D or L amino acid analog with a side chain consisting of the linker or bond to the bead. In one embodiment, the pseudo-free C-terminal residue may be of the D or the L optical configuration; in another embodiment, a racemic mixture of D and L-isomers may
25 be used.

In an additional embodiment, pyroglutamate may be included as the N-terminal residue of the peptide. Although pyroglutamate is not amenable to sequence by Edman degradation, by limiting substitution to only 50% of the peptides on a given bead with
30 N-terminal pyroglutamate, there will remain enough non-pyroglutamate peptide on the bead for sequencing. One of ordinary skill would readily recognize that this technique

could be used for sequencing of any peptide that incorporates a residue resistant to Edman degradation at the N-terminus. Other methods to characterize individual peptides that demonstrate desired activity are described in detail *infra*. Specific activity of a peptide that comprises a blocked N-terminal group, *e.g.*, pyroglutamate, when the particular N-terminal group is present in 50% of the peptides, would readily be demonstrated by comparing activity of a completely (100%) blocked peptide with a non-blocked (0%) peptide.

In addition, the present invention envisions preparing peptides that have more well defined structural properties, and the use of peptidomimetics, and peptidomimetic bonds, such as ester bonds, to prepare peptides with novel properties. In another embodiment, a peptide may be generated that incorporates a reduced peptide bond, *i.e.*, $R_1-CH_2-NH-R_2$, where R_1 and R_2 are amino acid residues or sequences. A reduced peptide bond may be introduced as a dipeptide subunit. Such a molecule would be resistant to peptide bond hydrolysis, *e.g.*, protease activity. Such peptides would provide ligands with unique function and activity, such as extended half-lives *in vivo* due to resistance to metabolic breakdown, or protease activity. Furthermore, it is well known that in certain systems constrained peptides show enhanced functional activity (Hruby, 1982, Life Sciences 31:189-199; Hruby et al., 1990, Biochem J. 268:249-262); the present invention provides a method to produce a constrained peptide that incorporates random sequences at all other positions.

A constrained, cyclic or rigidized peptide may be prepared synthetically, provided that in at least two positions in the sequence of the peptide an amino acid or amino acid analog is inserted that provides a chemical functional group capable of cross-linking to constrain, cyclise or rigidize the peptide after treatment to form the cross-link. Cyclization will be favored when a turn-inducing amino acid is incorporated. Examples of amino acids capable of cross-linking a peptide are cysteine to form disulfide, aspartic acid to form a lactone or a lactase, and a chelator such as γ -carboxyl-glutamic acid (Gla) (Bachem) to chelate a transition metal and form a cross-link. Protected γ -carboxyl glutamic acid may be prepared by modifying the synthesis described by Zee-Cheng and

Olson (1980, Biophys. Biochem. Res. Commun. 94:1128-1132). A peptide in which the peptide sequence comprises at least two amino acids capable of cross-linking may be treated, *e.g.*, by oxidation of cysteine residues to form a disulfide or addition of a metal ion to form a chelate, so as to cross-link the peptide and form a constrained, cyclic or
5 rigidized peptide.

The present invention provides strategies to systematically prepare cross-links. For example, if four cysteine residues are incorporated in the peptide sequence, different protecting groups may be used (Hiskey, 1981, in *The Peptides: Analysis, Synthesis, Biology*, Vol. 3, Gross and Meienhofer, eds., Academic Press: New York, pp. 137-167;
10 Ponsanti et al., 1990, *Tetrahedron* 46:8255-8266). The first pair of cysteine may be deprotected and oxidized, then the second set may be deprotected and oxidized. In this way a defined set of disulfide cross-links may be formed. Alternatively, a pair of cysteine and a pair of collating amino acid analogs may be incorporated so that the cross-
15 links are of a different chemical nature.

The following non-classical amino acids may be incorporated in the peptide in order to introduce particular conformational motifs: 1,2,3,4-tetrahydroisoquinoline-3-carboxylate (Kazmierski et al., 1991, *J. Am. Chem. Soc.* 113:2275-2283); (2S,3S)-methyl-
20 phenylalanine, (2S,3R)-methyl-phenylalanine, (2R,3S)-methyl-phenylalanine and (2R,3R)-methyl-phenylalanine (Kazmierski and Hruby, 1991, *Tetrahedron Lett.*); 2-aminotetrahydronaphthalene-2-carboxylic acid (Landis, 1989, Ph.D. Thesis, University of Arizona); hydroxy-1,2,3,4-tetrahydroisoquinoline-3-carboxylate (Miyake et al., 1989, *J. Takeda Res. Labs.* 43:53-76); β -carboline (D and L) (Kazmierski, 1988, Ph.D. Thesis,
25 University of Arizona); HIC (histidine isoquinoline carboxylic acid) (Zechel et al., 1991, *Int. J. Pep. Protein Res.* 43); and HIC (histidine cyclic urea) (Dharanipragada).

The following amino acid analogs and peptidomimetics may be incorporated into a peptide to induce or favor specific secondary structures: LL-Acp (LL-3-amino-
30 2-propenidone-6-carboxylic acid), a β -turn inducing dipeptide analog (Kemp et al., 1985,

J. Org. Chem. 50:5834-5838); β -sheet inducing analogs (Kemp et al., 1988, Tetrahedron Lett. 29:5081-5082); β -turn inducing analogs (Kemp et al., 1988, Tetrahedron Lett. 29:5057-5060); α -helix inducing analogs (Kemp et al., 1988, Tetrahedron Lett. 29:4935-4938); γ -turn inducing analogs (Kemp et al., 1989, J. Org. Chem. 54:109:115); and
5 analogs provided by the following references: Nagai and Sato, 1985, Tetrahedron Lett. 26:647-650; DiMaio et al., 1989, J. Chem. Soc. Perkin Trans. p. 1687; also a Gly-Ala turn analog (Kahn et al., 1989, Tetrahedron Lett. 30:2317); amide bond isostere (Jones et al., 1988, Tetrahedron Lett. 29:3853-3856); tetrazol (Zabrocki et al., 1988, J. Am. Chem. Soc. 110:5875-5880); DTC (Samanen et al., 1990, Int. J. Protein Pep. Res.
10 35:501:509); and analogs taught in Olson et al., 1990, J. Am. Chem. Sci. 112:323-333 and Garvey et al., 1990, J. Org. Chem. 56:436. Conformationally restricted mimetics of beta turns and beta bulges, and peptides containing them, are described in U.S. Patent No. 5,440,013, issued August 8, 1995 to Kahn.

15 The present invention further provides for modification or derivatization of the polypeptide or peptide of the invention. Modifications of peptides are well known to one of ordinary skill, and include phosphorylation, carboxymethylation, and acylation. Modifications may be effected by chemical or enzymatic means. In another aspect, glycosylated or fatty acylated peptide derivatives may be prepared. Preparation of
20 glycosylated or fatty acylated peptides is well known in the art. Fatty acyl peptide derivatives may also be prepared. For example, and not by way of limitation, a free amino group (N-terminal or lysyl) may be acylated, e.g., myristoylated. In another embodiment an amino acid comprising an aliphatic side chain of the structure - $(CH_2)_nCH_3$ may be incorporated in the peptide. This and other peptide-fatty acid
25 conjugates suitable for use in the present invention are disclosed in U.K. Patent GB-8809162.4, International Patent Application PCT/AU89/00166, and reference 5, *supra*.

Mutations can be made in a nucleic acid encoding the polypeptide such that a particular codon is changed to a codon which codes for a different amino acid. Such a mutation is
30 generally made by making the fewest nucleotide changes possible. A substitution mutation of this sort can be made to change an amino acid in the resulting protein in a

non-conservative manner (i.e., by changing the codon from an amino acid belonging to a grouping of amino acids having a particular size or characteristic to an amino acid belonging to another grouping) or in a conservative manner (i.e., by changing the codon from an amino acid belonging to a grouping of amino acids having a particular size or characteristic to an amino acid belonging to the same grouping). Such a conservative change generally leads to less change in the structure and function of the resulting protein. A non-conservative change is more likely to alter the structure, activity or function of the resulting protein. The present invention should be considered to include sequences containing conservative changes which do not significantly alter the activity or binding characteristics of the resulting protein. Substitutes for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. Amino acids containing aromatic ring structures are phenylalanine, tryptophan, and tyrosine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid. Such alterations will not be expected to affect apparent molecular weight as determined by polyacrylamide gel electrophoresis, or isoelectric point.

20

Particularly preferred substitutions are:

- Lys for Arg and vice versa such that a positive charge may be maintained;
- Glu for Asp and vice versa such that a negative charge may be maintained;
- Ser for Thr such that a free -OH can be maintained; and
- 25 - Gln for Asn such that a free NH_2 can be maintained.

Synthetic DNA sequences allow convenient construction of genes which will express analogs or "muteins". A general method for site-specific incorporation of unnatural amino acids into proteins is described in Noren, et al. *Science*, 244:182-188 (April 1989).

30 This method may be used to create analogs with unnatural amino acids.

In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Sambrook et al, "Molecular Cloning: A Laboratory Manual" (1989); "Current Protocols in Molecular Biology" Volumes I-III [Ausubel, R. M., ed. (1994)]; "Cell Biology: A Laboratory Handbook" Volumes I-III [J. E. Celis, ed. (1994)]; "Current Protocols in Immunology" Volumes I-III [Coligan, J. E., ed. (1994)]; "Oligonucleotide Synthesis" (M.J. Gait ed. 1984); "Nucleic Acid Hybridization" [B.D. Hames & S.J. Higgins eds. (1985)]; "Transcription And Translation" [B.D. Hames & S.J. Higgins, eds. (1984)]; "Animal Cell Culture" [R.I. Freshney, ed. (1986)]; "Immobilized Cells And Enzymes" [IRL Press, (1986)]; B. Perbal, "A Practical Guide To Molecular Cloning" (1984).

In an additional embodiment, pyroglutamate may be included as the N-terminal residue of the peptide. Although pyroglutamate is not amenable to sequence by Edman degradation, by limiting substitution to only 50% of the peptides on a given bead with N-terminal pyroglutamate, there will remain enough non-pyroglutamate peptide on the bead for sequencing. One of ordinary skill in the art would readily recognize that this technique could be used for sequencing of any peptide that incorporates a residue resistant to Edman degradation at the N-terminus. Other methods to characterize individual peptides that demonstrate desired activity are described in detail *infra*. Specific activity of a peptide that comprises a blocked N-terminal group, e.g., pyroglutamate, when the particular N-terminal group is present in 50% of the peptides, would readily be demonstrated by comparing activity of a completely (100%) blocked peptide with a non-blocked (0%) peptide.

25

Chemical Moieties For Derivatization. Chemical moieties suitable for derivatization may be selected from among water soluble polymers. The polymer selected should be water soluble so that the component to which it is attached does not precipitate in an aqueous environment, such as a physiological environment. Preferably, for therapeutic use of the end-product preparation, the polymer will be pharmaceutically acceptable. One skilled in the art will be able to select the desired polymer based on such

30

considerations as whether the polymer/component conjugate will be used therapeutically, and if so, the desired dosage, circulation time, resistance to proteolysis, and other considerations. For the present component or components, these may be ascertained using the assays provided herein.

- 5 The water soluble polymer may be selected from the group consisting of, for example, polyethylene glycol, copolymers of ethylene glycol/propylene glycol, carboxymethylcellulose, dextran, polyvinyl alcohol, polyvinyl pyrrolidone, poly-1, 3-dioxolane, poly-1,3,6-trioxane, ethylene/maleic anhydride copolymer, polyaminoacids (either homopolymers or random copolymers), and dextran or poly(n-vinyl
- 10 pyrrolidone)polyethylene glycol, propylene glycol homopolymers, polypropylene oxide/ethylene oxide co- polymers, polyoxyethylated polyols and polyvinyl alcohol. Polyethylene glycol propionaldehyde may have advantages in manufacturing due to its stability in water.
- 15 The number of polymer molecules so attached may vary, and one skilled in the art will be able to ascertain the effect on function. One may mono-derivatize, or may provide for a di-, tri-, tetra- or some combination of derivatization, with the same or different chemical moieties (*e.g.*, polymers, such as different weights of polyethylene glycols). The proportion of polymer molecules to component or components molecules will vary,
- 20 as will their concentrations in the reaction mixture. In general, the optimum ratio (in terms of efficiency of reaction in that there is no excess unreacted component or components and polymer) will be determined by factors such as the desired degree of derivatization (*e.g.*, mono, di-, tri-, etc.), the molecular weight of the polymer selected, whether the polymer is branched or unbranched, and the reaction conditions.
- 25 The polyethylene glycol molecules (or other chemical moieties) should be attached to the component or components with consideration of effects on functional or antigenic domains of the protein. There are a number of attachment methods available to those skilled in the art, *e.g.*, EP 0 401 384 herein incorporated by reference (coupling PEG to
- 30 G-CSF), *see also* Malik et al., 1992, Exp. Hematol. 20:1028-1035 (reporting pegylation of GM-CSF using tresyl chloride). For example, polyethylene glycol may be covalently

bound through amino acid residues via a reactive group, such as, a free amino or carboxyl group. Reactive groups are those to which an activated polyethylene glycol molecule may be bound. The amino acid residues having a free amino group include lysine residues and the - terminal amino acid residues; those having a free carboxyl group
5 include aspartic acid residues glutamic acid residues and the C-terminal amino acid residue. Sulfhydryl groups may also be used as a reactive group for attaching the polyethylene glycol molecule(s). Preferred for therapeutic purposes is attachment at an amino group, such as attachment at the N-terminus or lysine group.

- 10 This invention provides a method for determining whether a subject carries a mutation in the SH3D1A gene which comprises: a) obtaining an appropriate nucleic acid sample from the subject; and (b) determining whether the nucleic acid sample from step (a) is, or is derived from, a nucleic acid which encodes mutant SH3D1A so as to thereby determine whether a subject carries a mutation in the SH3D1A gene. In one embodiment,
15 the nucleic acid sample in step (a) comprises mRNA corresponding to the transcript of DNA encoding a mutant SH3D1A, and wherein the determining of step (b) comprises: (i) contacting the mRNA with the oligonucleotide under conditions permitting binding of the mRNA to the oligonucleotide so as to form a complex; (ii) isolating the complex so formed; and (iii) identifying the mRNA in the isolated complex so as to thereby
20 determine whether the mRNA is, or is derived from, a nucleic acid which encodes mutant SH3D1A. In another embodiment, the determining of step (b) comprises: i) contacting the nucleic acid sample of step (a), and the isolated nucleic acid with restriction enzymes under conditions permitting the digestion of the nucleic acid sample, and the isolated nucleic acid into distinct, distinguishable pieces of nucleic acid; (ii) isolating the pieces
25 of nucleic acid; and (iii) comparing the pieces of nucleic acid derived from the nucleic acid sample with the pieces of nucleic acid derived from the isolated nucleic acid so as to thereby determine whether the nucleic acid sample is, or is derived from, a nucleic acid which encodes mutant SH3D1A.
- 30 The present invention further provides methods of preparing a polynucleotide comprising polymerizing nucleotides to yield a sequence comprised of at least eight

consecutive nucleotides of the SH3D1A gene; and methods of preparing a polypeptide comprising polymerizing amino acids to yield a sequence comprising at least five amino acids encoded within the SH3D1A gene.

- 5 The present invention further provides methods of screening the SH3D1A gene to identify mutations. Such methods may further comprise the step of amplifying a portion of the SH3D1A gene, and may further include a step of providing a set of polynucleotides which are primers for amplification of said portion of the SH3D1A gene. The method is useful for identifying mutations for use in either diagnosis of the
- 10 predisposition to, and diagnosis and treatment of megakaryocytic abnormality, hematopoietic disorders, myeloproliferative disorder, platelet disorder, leukemia; neural abnormality or other disorder; and prenatal diagnosis and treatment of tumors. Useful diagnostic techniques include, but are not limited to fluorescent in situ hybridization (FISH), direct DNA sequencing, PFGE analysis, Southern blot analysis, single
- 15 stranded conformation analysis (SSCA), Rnase protection assay, allele-specific oligonucleotide (ASO), dot blot analysis and PCR-SSCP, as discussed in detail further below.

- There are several methods that can be used to detect DNA sequence variation.
- 20 Direct DNA sequencing, either manual sequencing or automated fluorescent sequencing can detect sequence variation. For a gene as large as SH3D1A, manual sequencing is very labor-intensive, but under optimal conditions, mutations in the coding sequence of a gene are rarely missed. Another approach is the single-stranded conformation polymorphism assay (SSCA) (Orita et al., 1989). This method does not
- 25 detect all sequence changes, especially if the DNA fragment size is greater than 200 bp, but can be optimized to detect most DNA sequence variation. The reduced detection sensitivity is a disadvantage, but the increased throughput possible with SSCA makes it an attractive, viable alternative to direct sequencing for mutation detection on a research basis. The fragments which have shifted mobility on SSCA gels are then
- 30 sequenced to determine the exact nature of the DNA sequence variation. Other approaches based on the detection of mismatches between the two complementary DNA

strands include clamped denaturing gel electrophoresis (CDGE) (Sheffield et al., 1991), heteroduplex analysis (HA) (White et al., 1992) and chemical mismatch cleavage (CMC) (Grompe et al., 1989). None of the methods described above will detect large deletions, duplications or insertions, nor will they detect a regulatory mutation which affects transcription or translation of the protein. Other methods which might detect these classes of mutations such as a protein truncation assay or the asymmetric assay, detect only specific types of mutations and would not detect missense mutations. A review of currently available methods of detecting DNA sequence variation can be found in a recent review by Grompe (1993). Once a mutation is known, an allele specific detection approach such as allele specific oligonucleotide (ASO) hybridization can be utilized to rapidly screen large numbers of other samples for that same mutation.

A rapid preliminary analysis to detect polymorphisms in DNA sequences can be performed by looking at a series of Southern blots of DNA cut with one or more restriction enzymes, preferably with a large number of restriction enzymes. Each blot contains a series of normal individuals and a series of tumors. Southern blots displaying hybridizing fragments (differing in length from control DNA when probed with sequences near or including the SH3D1A gene) indicate a possible mutation. If restriction enzymes which produce very large restriction fragments are used, then pulsed field gel electrophoresis (PFGE) is employed.

Detection of point mutations may be accomplished by molecular cloning of the SH3D1A allele(s) and sequencing the allele(s) using techniques well known in the art. Alternatively, the gene sequences can be amplified directly from a genomic DNA preparation from the tumor tissue, using known techniques. The DNA sequence of the amplified sequences can then be determined. There are six well known methods for a more complete, yet still indirect, test for confirming the presence of a susceptibility allele: 1) single stranded conformation analysis (SSCA) (Orita et al., 1989); 2) denaturing gradient gel electrophoresis (DGGE) (Wartell et al., 1990; Sheffield et al., 1989); 3) RNase protection assays (Finkelstein et al., 1990; Kinszler et al., 1991); 4)

of allele-specific probes with amplified SH3D1A sequences can be performed, for example, on a nylon filter. Hybridization to a particular probe under stringent hybridization conditions indicates the presence of the same mutation in the tumor tissue as in the allele-specific probe.

5

Alteration of SH3D1A mRNA expression can be detected by any techniques known in the art. These include Northern blot analysis, PCR amplification and RNase protection. Diminished mRNA expression indicates an alteration of the wild-type SH3D1A gene. Alteration of wild-type SH3D1A genes can also be detected by
10 screening for alteration of wild-type SH3D1A protein. For example, monoclonal antibodies immunoreactive with SH3D1A can be used to screen a tissue. Lack of cognate antigen would indicate a SH3D1A mutation. Antibodies specific for products of mutant alleles could also be used to detect mutant SH3D1A gene product. Such immunological assays can be done in any convenient formats known in the art. These
15 include Western blots, immunohistochemical assays and ELISA assays. Any means for detecting an altered SH3D1A protein can be used to detect alteration of wild-type SH3D1A genes. Functional assays, such as protein binding determinations, can be used. In addition, assays can be used which detect SH3D1A biochemical function. Finding a mutant SH3D1A gene product indicates alteration of a wild-type SH3D1A
20 gene. Mutant SH3D1A genes or gene products can also be detected in other human body samples, such as serum, stool, urine and sputum.

The present invention also provides for fusion polypeptides, comprising SH3D1A polypeptides and fragments. Homologous polypeptides may be fusions between two or
25 more SH3D1A polypeptide sequences or between the sequences of SH3D1A and a related protein. Likewise, heterologous fusions may be constructed which would exhibit a combination of properties or activities of the derivative proteins. For example, ligand-binding or other domains may be "swapped" between different new fusion polypeptides or fragments. Such homologous or heterologous fusion polypeptides
30 may display, for example, altered strength or specificity of binding. Fusion partners include immunoglobulins, bacterial beta -galactosidase, trpE, protein A, beta

allele-specific oligonucleotides (ASOs) (Conner et al., 1983); 5) the use of proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein (Modrich, 1991); and 6) allele-specific PCR (Rano & Kidd, 1989). For allele-specific PCR, primers are used which hybridize at their 3' ends to a particular SH3D1A mutation.

5 If the particular SH3D1A mutation is not present, an amplification product is not observed. Amplification Refractory Mutation System (ARMS) can also be used, as disclosed in European Patent Application Publication No. 0332435 and in Newton et al., 1989. Insertions and deletions of genes can also be detected by cloning, sequencing and amplification. In addition, restriction fragment length polymorphism (RFLP) probes for
10 the gene or surrounding marker genes can be used to score alteration of an allele or an insertion in a polymorphic fragment. Such a method is particularly useful for screening relatives of an affected individual for the presence of the SH3D1A mutation found in that individual. Other techniques for detecting insertions and deletions as known in the art can be used.

15 In similar fashion, DNA probes can be used to detect mismatches, through enzymatic or chemical cleavage. See, e.g., Cotton et al., 1988; Shenk et al., 1975; Novack et al., 1986. Alternatively, mismatches can be detected by shifts in the electrophoretic mobility of mismatched duplexes relative to matched duplexes. See, e.g., Cariello, 1988. With
20 either riboprobes or DNA probes, the cellular mRNA or DNA which might contain a mutation can be amplified using PCR (see below) before hybridization. Changes in DNA of the SH3D1A gene can also be detected using Southern hybridization, especially if the changes are gross rearrangements, such as deletions and insertions.

25 DNA sequences of the SH3D1A gene which have been amplified by use of PCR may also be screened using allele-specific probes. These probes are nucleic acid oligomers, each of which contains a region of the SH3D1A gene sequence harboring a known mutation. For example, one oligomer may be about 30 nucleotides in length, corresponding to a portion of the SH3D1A gene sequence. By use of a battery of
30 such allele-specific probes, PCR amplification products can be screened to identify the presence of a previously identified mutation in the SH3D1A gene. Hybridization

-lactamase, alpha amylase, alcohol dehydrogenase and yeast alpha mating factor. See, e.g., Godowski et al. , 1988. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described below, or may be chemically synthesized. Techniques for the synthesis of polypeptides are described, for example, in Merrifield, 1963.

This invention provides a method for determining whether a subject has a megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, or leukemia which comprises: (a) obtaining an appropriate sample from the subject; and (b) contacting the sample with the antibody so as to thereby determine whether a subject has the megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, or leukemia.

This invention provides a method for determining whether a subject has a predisposition for a megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, leukemia or a neural abnormality or other disorder, which comprises: (a) obtaining an appropriate nucleic acid sample from the subject; and (b) determining whether the nucleic acid sample from step (a) is, or is derived from, a nucleic acid which encodes SH3D1A so as to thereby determine whether a subject has a predisposition for a megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, or leukemia.

This invention provides a method for determining whether a subject has a megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, leukemia or a neural abnormality or other disorder, which comprises: (a) obtaining an appropriate nucleic acid sample from the subject; and (b) determining whether the nucleic acid sample from step (a) is, or is derived from, a nucleic acid which encodes the human SH3D1A so as to thereby determine whether a subject has megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, leukemia or a neural abnormality or other disorder. In one embodiment the nucleic acid sample in step (a) comprises mRNA corresponding to the transcript of DNA encoding a human SH3D1A, and wherein the determining of step (b) comprises: (i) contacting the mRNA with the oligonucleotide

under conditions permitting binding of the mRNA to the oligonucleotide so as to form a complex; (ii) isolating the complex so formed; and (iii) identifying the mRNA in the isolated complex so as to thereby determine whether the mRNA is, or is derived from, a nucleic acid which encodes a human SH3D1A. A particular finding in accordance with the invention, is that such disorders as may occur in adult brain have been observed with respect to the present invention, and accordingly adult patients may be diagnosed, and if possible, treated by the application of the inventive subject matter hereof.

This invention provides a method of suppressing cells unable to regulate themselves which comprises introducing a purified human SH3D1A into the cells in an amount effective to suppress the cells.

This invention provides a method for identifying a chemical compound which is capable of suppressing cells unable to regulate themselves in a subject which comprises: (a) contacting the SH3D1A with a chemical compound under conditions permitting binding between the SH3D1A and the chemical compound; (b) detecting specific binding of the chemical compound to the SH3D1A; and (c) determining whether the chemical compound inhibits the SH3D1A so as to identify a chemical compound which is capable of suppressing cells unable to regulate themselves.

20

This invention provides a method for screening a tumor sample from a human subject for a somatic alteration in a SH3D1A gene in said tumor which comprises gene comparing a first sequence selected from the group consisting of a SH3D1A gene from said tumor sample, SH3D1A RNA from said tumor sample and SH3D1A cDNA made from mRNA from said tumor sample with a second sequence selected from the group consisting of SH3D1A gene from a nontumor sample of said subject, SH3D1A RNA from said nontumor sample and SH3D1A cDNA made from mRNA from said nontumor sample, wherein a difference in the sequence of the SH3D1A gene, SH3D1A RNA or SH3D1A cDNA from said tumor sample from the sequence of the SH3D1A gene, SH3D1A RNA or SH3D1A cDNA from said nontumor sample indicates a somatic alteration in the SH3D1A gene in said tumor sample.

This invention provides a method for screening a tumor sample from a human subject for the presence of a somatic alteration in a SH3D1A gene in said tumor which comprises comparing SH3D1A polypeptide from said tumor sample from said subject to SH3D1A polypeptide from a nontumor sample from said subject to analyze for a difference
5 between the polypeptides, wherein said comparing is performed by (i) detecting either a full length polypeptide or a truncated polypeptide in each sample or (ii) contacting an antibody which specifically binds to either an epitope of an altered SH3D1A polypeptide or an epitope of a wild-type SH3D1A polypeptide to the SH3D1A polypeptide from each sample and detecting antibody binding, wherein a difference between the SH3D1A
10 polypeptide from said tumor sample from the SH3D1A polypeptide from said nontumor sample indicates the presence of a somatic alteration in the SH3D1A gene in said tumor sample.

This invention provides a method for monitoring the progress and adequacy of treatment
15 in a subject who has received treatment for a megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, leukemia or a condition involving a neural abnormality or dysfunction, which comprises monitoring the level of nucleic acid encoding the human SH3D1A at various stages of treatment.

20 This invention provides a pharmaceutical composition comprising an amount of a polypeptide of the present invention, and a pharmaceutically effective carrier or diluent.

This invention provides a method of treating a subject having megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, or leukemia which comprises
25 introducing the isolated nucleic acid into the subject under conditions such that the nucleic acid expresses SH3D1A, so as to thereby treat the subject.

This invention provides a method of treating a subject having megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, or leukemia which comprises
30 administration to the subject a therapeutically effective amount of the pharmaceutical composition to the subject.

This invention is directed to diagnostic methods and therepeutic treatments relating to th e following: Wilms tumor, Li-Fraumcini syndrome, retinoblastoma, familiar colon cancer, and acute myelogenous leukemia (AML), and myelodysplastic syndromes (MDSs).

5

Further, it is contemplated by this invention that the disclosed invention is directed to diversified hereditary disorders of platelet production. Heredity disorders of platelet production include but is not limited to: clinical problems in these disorders range from mild cutaneous petechiae or occasional epistaxes to severe hemorrhage requiring red cell and platelet transfusions; and abnormalities of thrombocyte structure, function, and number have been found by laboratory evaluation of some of these patients. Deviations from normality in various components of the platelet response during hemostatis have been well characterized in a number of families and are known to those skilled in the art. These include defects of platelet adhesion, secretion from storage granules, and subsequent aggregation.

15

This invention provides a method of diagnosing megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, or leukemia in a subject which comprises: (a) obtaining a nucleic acid molecule from a tumor lesion of the subject; (b) contacting the nucleic acid molecule with a labelled nucleic acid molecule of at least 15 nucleotides capable of specifically hybridizing with the isolated DNA, under hybridizing conditions; and (c) determining the presence of the nucleic acid molecule hybridized, the presence of which is indicative of megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, or leukemia in the subject, thereby diagnosing megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, or leukemia in the subject.

25

In one embodiment the DNA molecule from the tumor lesion is amplified before step (b). In another embodiment PCR is employed to amplify the nucleic acid molecule. Methods of amplifying nucleic acid molecules are known to those skilled in the art.

30

In the above described methods, a size fractionation may be employed which is effected by a polyacrylamide gel. In one embodiment, the size fractionation is effected by an agarose gel. Further, transferring the DNA fragments into a solid matrix may be employed before a hybridization step. One example of such solid matrix is nitrocellulose paper.

This invention provides a method of diagnosing megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, leukemia or a neural abnormality or dysfunction, in a subject which comprises: (a) obtaining a nucleic acid molecule from a suitable bodily fluid of the subject; (b) contacting the nucleic acid molecule with a labelled nucleic acid molecules of at least 15 nucleotides capable of specifically hybridizing with the isolated DNA, under hybridizing conditions; and (c) determining the presence of the nucleic acid molecule hybridized, the presence of which is indicative of megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, leukemia or neural abnormality or dysfunction, in the subject, thereby diagnosing megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, or leukemia in the subject.

This invention provides a method of diagnosing a DNA virus in a subject, which comprises (a) obtaining a suitable bodily fluid sample from the subject, (b) contacting the suitable bodily fluid of the subject to a support having already bound thereto a antibody, so as to bind the antibody to a specific antigen, (c) removing unbound bodily fluid from the support, and (d) determining the level of antibody bound by the antigen, thereby diagnosing the subject for megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, leukemia or neural disorder.

This invention provides a method of diagnosing megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, or leukemia in a subject, which comprises (a) obtaining a suitable bodily fluid sample from the subject, (b) contacting the suitable bodily fluid of the subject to a support having already bound thereto an antigen, so as to bind antigen to a specific antibody, (c) removing unbound bodily fluid from the support,

and (d) determining the level of the antigen bound by the antibody, thereby diagnosing megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, leukemia or neural disorder.

- 5 A suitable bodily fluid includes, but is not limited to: serum, plasma, cerebrospinal fluid, lymphocytes, urine, transudates, or exudates. In the preferred embodiment, the suitable bodily fluid sample is serum or plasma. In addition, the bodily fluid sample may be cells from bone marrow, or a supernatant from a cell culture. Methods of obtaining a suitable bodily fluid sample from a subject are known to those skilled in the art. Methods of
- 10 determining the level of antibody or antigen include, but are not limited to: ELISA, IFA, and Western blotting.

- The diagnostic assays of the invention can be nucleic acid assays such as nucleic acid hybridization assays and assays which detect amplification of specific nucleic acid to
- 15 detect for a nucleic acid sequence of the human SH3D1A described herein.

- Accepted means for conducting hybridization assays are known and general overviews of the technology can be had from a review of: *Nucleic Acid Hybridization: A Practical Approach* [72]; *Hybridization of Nucleic Acids Immobilized on Solid Supports* [41];
- 20 *Analytical Biochemistry* [4] and Innis *et al.*, *PCR Protocols* [74], *supra*, all of which are incorporated by reference herein.

- Target specific probes may be used in the nucleic acid hybridization diagnostic. The probes are specific for or complementary to the target of interest. For precise allelic
- 25 differentiations, the probes should be about 14 nucleotides long and preferably about 20-30 nucleotides. For more general detection of the human SH3D1A of the invention, nucleic acid probes are about 50 to about 1000 nucleotides, most preferably about 200 to about 400 nucleotides.

- 30 The specific nucleic acid probe can be RNA or DNA polynucleotide or oligonucleotide, or their analogs. The probes may be single or double stranded nucleotides. The probes

of the invention may be synthesized enzymatically, using methods well known in the art (e.g., nick translation, primer extension, reverse transcription, the polymerase chain reaction, and others) or chemically (e.g., by methods such as the phosphoramidite method described by Beaucage and Carruthers [19], or by the triester method according to

5 Matteucci, *et al.* [62], both incorporated herein by reference).

An alternative means for determining the presence of the human SH3D1A is *in situ* hybridization, or more recently, *in situ* polymerase chain reaction. *In situ* PCR is described in Neuvo *et al.* [71], Intracellular localization of polymerase chain reaction

10 (PCR)-amplified Hepatitis C cDNA; Bagasra *et al.* [10], Detection of Human Immunodeficiency virus type 1 provirus in mononuclear cells by *in situ* polymerase chain reaction; and Heniford *et al.* [35], Variation in cellular EGF receptor mRNA expression demonstrated by *in situ* reverse transcriptase polymerase chain reaction. *In situ* hybridization assays are well known and are generally described in *Methods Enzymol.*

15 [67] incorporated by reference herein. In an *in situ* hybridization, cells are fixed to a solid support, typically a glass slide. The cells are then contacted with a hybridization solution at a moderate temperature to permit annealing of target-specific probes that are labeled. The probes are preferably labelled with radioisotopes or fluorescent reporters.

20 The above described probes are also useful for in-situ hybridization or in order to locate tissues which express this gene, or for other hybridization assays for the presence of this gene or its mRNA in various biological tissues. In-situ hybridization is a sensitive localization method which is not dependent on expression of antigens or native vs. denatured conditions.

25 In brief, inhibitory nucleic acid therapy approaches can be classified into those that target DNA sequences, those that target RNA sequences (including pre-mRNA and mRNA), those that target proteins (sense strand approaches), and those that cause cleavage or chemical modification of the target nucleic acids.

30 Approaches targeting DNA fall into several categories. Nucleic acids can be designed to bind to the major groove of the duplex DNA to form a triple helical or "triplex"

structure. Alternatively, inhibitory nucleic acids are designed to bind to regions of single stranded DNA resulting from the opening of the duplex DNA during replication or transcription.

5 More commonly, inhibitory nucleic acids are designed to bind to mRNA or mRNA precursors. Inhibitory nucleic acids are used to prevent maturation of pre-mRNA. Inhibitory nucleic acids may be designed to interfere with RNA processing, splicing or translation.

10 The inhibitory nucleic acids can be targeted to mRNA. In this approach, the inhibitory nucleic acids are designed to specifically block translation of the encoded protein. Using this approach, the inhibitory nucleic acid can be used to selectively suppress certain cellular functions by inhibition of translation of mRNA encoding critical proteins. For example, an inhibitory nucleic acid complementary to regions of c-myc mRNA inhibits
15 c-myc protein expression in a human promyelocytic leukemia cell line, HL60, which overexpresses the c-myc proto-oncogene. See Wickstrom E.L., *et al.* [93] and Harel-Bellan, A., *et al.* [31A]. As described in Helene and Toulme, inhibitory nucleic acids targeting mRNA have been shown to work by several different mechanisms to inhibit translation of the encoded protein(s).

20

Lastly, the inhibitory nucleic acids can be used to induce chemical inactivation or cleavage of the target genes or mRNA. Chemical inactivation can occur by the induction of crosslinks between the inhibitory nucleic acid and the target nucleic acid within the cell. Other chemical modifications of the target nucleic acids induced by appropriately
25 derivatized inhibitory nucleic acids may also be used.

Cleavage, and therefore inactivation, of the target nucleic acids may be effected by attaching a substituent to the inhibitory nucleic acid which can be activated to induce cleavage reactions. The substituent can be one that affects either chemical, or enzymatic
30 cleavage. Alternatively, cleavage can be induced by the use of ribozymes or catalytic RNA. In this approach, the inhibitory nucleic acids would comprise either naturally

occurring RNA (ribozymes) or synthetic nucleic acids with catalytic activity.

used herein, "pharmaceutical composition" could mean therapeutically effective amounts of polypeptide products of the invention together with suitable diluents, preservatives, solubilizers, emulsifiers, adjuvant and/or carriers useful in SCF (stem cell factor) therapy. A "therapeutically effective amount" as used herein refers to that amount which provides a therapeutic effect for a given condition and administration regimen. Such compositions are liquids or lyophilized or otherwise dried formulations and include diluents of various buffer content (e.g., Tris-HCl, acetate, phosphate), pH and ionic strength, additives such as albumin or gelatin to prevent absorption to surfaces, detergents (e.g., Tween 20, Tween 80, Pluronic F68, bile acid salts), solubilizing agents (e.g., glycerol, polyethylene glycerol), anti-oxidants (e.g., ascorbic acid, sodium metabisulfite), preservatives (e.g., Thimerosal, benzyl alcohol, parabens), bulking substances or tonicity modifiers (e.g., lactose, mannitol), covalent attachment of polymers such as polyethylene glycol to the protein, complexation with metal ions, or incorporation of the material into or onto particulate preparations of polymeric compounds such as polylactic acid, polglycolic acid, hydrogels, etc, or onto liposomes, microemulsions, micelles, unilamellar or multilamellar vesicles, erythrocyte ghosts, or spheroplasts. Such compositions will influence the physical state, solubility, stability, rate of in vivo release, and rate of *in vivo* clearance of SCF. The choice of compositions will depend on the physical and chemical properties of the protein having SCF activity. For example, a product derived from a membrane-bound form of SCF may require a formulation containing detergent. Controlled or sustained release compositions include formulation in lipophilic depots (e.g., fatty acids, waxes, oils). Also comprehended by the invention are particulate compositions coated with polymers (e.g., poloxamers or poloxamines) and SCF coupled to antibodies directed against tissue-specific receptors, ligands or antigens or coupled to ligands of tissue-specific receptors. Other embodiments of the compositions of the invention incorporate particulate forms protective coatings, protease inhibitors or permeation enhancers for various routes of administration, including parenteral, pulmonary, nasal and oral.

- Further, as used herein "pharmaceutically acceptable carrier" are well known to those skilled in the art and include, but are not limited to, 0.01-0.1M and preferably 0.05M phosphate buffer or 0.8% saline. Additionally, such pharmaceutically acceptable carriers may be aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of
- 5 non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's or fixed oils. Intravenous vehicles
- 10 include fluid and nutrient replenishers, electrolyte replenishers such as those based on Ringer's dextrose, and the like. Preservatives and other additives may also be present, such as, for example, antimicrobials, antioxidants, collating agents, inert gases and the like.
- 15 The term "adjuvant" refers to a compound or mixture that enhances the immune response to an antigen. An adjuvant can serve as a tissue depot that slowly releases the antigen and also as a lymphoid system activator that non-specifically enhances the immune response (Hood et al., *Immunology, Second Ed.*, 1984, Benjamin/Cummings: Menlo Park, California, p. 384). Often, a primary challenge with an antigen alone, in the
- 20 absence of an adjuvant, will fail to elicit a humoral or cellular immune response. Adjuvant include, but are not limited to, complete Freund's adjuvant, incomplete Freund's adjuvant, saponin, mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil or hydrocarbon emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful human
- 25 adjuvant such as BCG (*bacille Calmette-Guerin*) and *Corynebacterium parvum*. Preferably, the adjuvant is pharmaceutically acceptable.

- Controlled or sustained release compositions include formulation in lipophilic depots (e.g. fatty acids, waxes, oils). Also comprehended by the invention are particulate
- 30 compositions coated with polymers (e.g. poloxamers or poloxamines) and the compound coupled to antibodies directed against tissue-specific receptors, ligands or antigens or

coupled to ligands of tissue-specific receptors. Other embodiments of the compositions of the invention incorporate particulate forms protective coatings, protease inhibitors or permeation enhancers for various routes of administration, including parenteral, pulmonary, nasal and oral.

5

When administered, compounds are often cleared rapidly from mucosal surfaces or the circulation and may therefore elicit relatively short-lived pharmacological activity. Consequently, frequent administrations of relatively large doses of bioactive compounds may be required to sustain therapeutic efficacy. Compounds modified by the covalent attachment of water-soluble polymers such as polyethylene glycol, copolymers of polyethylene glycol and polypropylene glycol, carboxymethyl cellulose, dextran, polyvinyl alcohol, polyvinylpyrrolidone or polyproline are known to exhibit substantially longer half-lives in blood following intravenous injection than do the corresponding unmodified compounds (Abuchowski et al., 1981; Newmark et al., 1982; and Katre et al., 1987). Such modifications may also increase the compound's solubility in aqueous solution, eliminate aggregation, enhance the physical and chemical stability of the compound, and greatly reduce the immunogenicity and reactivity of the compound. As a result, the desired *in vivo* biological activity may be achieved by the administration of such polymer-compound adducts less frequently or in lower doses than with the unmodified compound.

Dosages. The sufficient amount may include but is not limited to from about 1 $\mu\text{g/kg}$ to about 1000 mg/kg. The amount may be 10 mg/kg. The pharmaceutically acceptable form of the composition includes a pharmaceutically acceptable carrier.

25

The preparation of therapeutic compositions which contain an active component is well understood in the art. Typically, such compositions are prepared as an aerosol of the polypeptide delivered to the nasopharynx or as injectables, either as liquid solutions or suspensions, however, solid forms suitable for solution in, or suspension in, liquid prior to injection can also be prepared. The preparation can also be emulsified. The active therapeutic ingredient is often mixed with excipients which are pharmaceutically

30

acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the composition can contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents which enhance the effectiveness of the active ingredient.

An active component can be formulated into the therapeutic composition as neutralized pharmaceutically acceptable salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with the free amino groups of the polypeptide or antibody molecule) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed from the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

A composition comprising "A" (where "A" is a single protein, DNA molecule, vector, etc.) is substantially free of "B" (where "B" comprises one or more contaminating proteins, DNA molecules, vectors, etc.) when at least about 75% by weight of the proteins, DNA, vectors (depending on the category of species to which A and B belong) in the composition is "A". Preferably, "A" comprises at least about 90% by weight of the A+B species in the composition, most preferably at least about 99% by weight.

The phrase "therapeutically effective amount" is used herein to mean an amount sufficient to reduce by at least about 15 percent, preferably by at least 50 percent, more preferably by at least 90 percent, and most preferably prevent, a clinically significant deficit in the activity, function and response of the host.

According to the invention, the component or components of a therapeutic composition of the invention may be introduced parenterally, transmucosally, *e.g.*, orally, nasally, pulmonarily, or rectally, or transdermally. Preferably, administration is parenteral, *e.g.*,

via intravenous injection, and also including, but is not limited to, intra-arteriole, intramuscular, intradermal, subcutaneous, intraperitoneal, intraventricular, and intracranial administration. Oral or pulmonary delivery may be preferred to activate mucosal immunity; since pneumococci generally colonize the nasopharyngeal and pulmonary mucosa, mucosal immunity may be a particularly effective preventive treatment. The term "unit dose" when used in reference to a therapeutic composition of the present invention refers to physically discrete units suitable as unitary dosage for humans, each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required diluent; *i.e.*, carrier, or vehicle.

In another embodiment, the active compound can be delivered in a vesicle, in particular a liposome (see Langer, *Science* 249:1527-1533 (1990); Treat et al., in *Liposomes in the Therapy of Infectious Disease and Cancer*, Lopez-Berestein and Fidler (eds.), Liss, New York, pp. 353-365 (1989); Lopez-Berestein, *ibid.*, pp. 317-327; see generally *ibid.*).

In yet another embodiment, the therapeutic compound can be delivered in a controlled release system. For example, the polypeptide may be administered using intravenous infusion, an implantable osmotic pump, a transdermal patch, liposomes, or other modes of administration. In one embodiment, a pump may be used (see Langer, *supra*; Sefton, *CRC Crit. Ref. Biomed. Eng.* 14:201 (1987); Buchwald et al., *Surgery* 88:507 (1980); Saudek et al., *N. Engl. J. Med.* 321:574 (1989)). In another embodiment, polymeric materials can be used (see *Medical Applications of Controlled Release*, Langer and Wise (eds.), CRC Pres., Boca Raton, Florida (1974); *Controlled Drug Bioavailability, Drug Product Design and Performance*, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, *J. Macromol. Sci. Rev. Macromol. Chem.* 23:61 (1983); see also Levy et al., *Science* 228:190 (1985); During et al., *Ann. Neurol.* 25:351 (1989); Howard et al., *J. Neurosurg.* 71:105 (1989)). In yet another embodiment, a controlled release system can be placed in proximity of the therapeutic target, *i.e.*, the brain, thus requiring only a fraction of the systemic dose (see, e.g., Goodson, in *Medical Applications of Controlled Release*, *supra*, vol. 2, pp. 115-138 (1984)). Preferably, a controlled release

device is introduced into a subject in proximity of the site of inappropriate immune activation or a tumor. Other controlled release systems are discussed in the review by Langer 1990, *Science* 249:1527-1533.

- 5 A subject in whom administration of an active component as set forth above is an effective therapeutic regimen for a bacterial infection is preferably a human, but can be any animal. Thus, as can be readily appreciated by one of ordinary skill in the art, the methods and pharmaceutical compositions of the present invention are particularly suited to administration to any animal, particularly a mammal, and including, but by no means
- 10 limited to, domestic animals, such as feline or canine subjects, farm animals, such as but not limited to bovine, equine, caprine, ovine, and porcine subjects, wild animals (whether in the wild or in a zoological garden), research animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats, etc., *i.e.*, for veterinary medical use.
- 15 In the therapeutic methods and compositions of the invention, a therapeutically effective dosage of the active component is provided. A therapeutically effective dosage can be determined by the ordinary skilled medical worker based on patient characteristics (age, weight, sex, condition, complications, other diseases, etc.), as is well known in the art. Furthermore, as further routine studies are conducted, more specific information will
- 20 emerge regarding appropriate dosage levels for treatment of various conditions in various patients, and the ordinary skilled worker, considering the therapeutic context, age and general health of the recipient, is able to ascertain proper dosing. Generally, for intravenous injection or infusion, dosage may be lower than for intraperitoneal, intramuscular, or other route of administration. The dosing schedule may vary,
- 25 depending on the circulation half-life, and the formulation used. The compositions are administered in a manner compatible with the dosage formulation in the therapeutically effective amount. Precise amounts of active ingredient required to be administered depend on the judgment of the practitioner and are peculiar to each individual. However, suitable dosages may range from about 0.1 to 20, preferably about 0.5 to about 10, and
- 30 more preferably one to several, milligrams of active ingredient per kilogram body weight of individual per day and depend on the route of administration. Suitable regimes for

initial administration and booster shots are also variable, but are typified by an initial administration followed by repeated doses at one or more hour intervals by a subsequent injection or other administration. Alternatively, continuous intravenous infusion sufficient to maintain concentrations of ten nanomolar to ten micromolar in the blood are
5 contemplated.

This invention is illustrated in the Experimental Details section which follows. These sections are set forth to aid in an understanding of the invention but are not intended to, and should not be construed to, limit in any way the invention as set forth in the claims
10 which follow thereafter.

EXPERIMENTAL DETAILS SECTION

The invention discloses a small candidate region of 50-200 kb for low platelets in
15 deletion for chromosome 21. At present, the candidate region for the familial platelet disorder is greater than 3,000 kb, a region containing as many as 150 genes. The SH3D1A is mapped to the small candidate region for low platelets for chromosome 21. Northern analysis using new sequence from SH3D1A reveals an abnormal band with significantly higher expression in RNA from lymphoblastoid cells derived from an
20 affected individual vs. normal controls. DNA sequence analyses reveal homologies to domains that suggest involvement in developmental and/or cell regulatory phenomena such as lead to cancers when disturbed. These include the SH3 domains as well as EH domains, both associated with protein-protein interactions and the latter associated with maintenance of the cytoskeleton. Therefore, mutations, or increased
25 or decreased expression are ultimately responsible for familial platelet disorder and possibly also for DS leukemias, subsets of non-DS leukemias and the processes that ultimately lead to abnormal platelets associated with deletion of chromosome 21.

Materials and Methods

30

Genomic clone obtained by screening the BAC library with EST: In order to study

the gene structure of SH3D1A, the genomic clones were obtained by screening a human BAC library B with a radio-labeled EST (cDNA) (dbEST#482496, Research Genetics, AL) according to the procedure described by Hurbet et al., 1997. Three positive clones were observed.

5

Fluorescence in situ hybridization (FISH) to confirm the cytogenetic location of BAC 119E16 on chromosomes 21q22,11-12: BAC DNAs were made as described in the previous publication (Hurbert et al., 1997). The BAC DNAs as probes were biotinylated and FISHed onto normal human chromosome preparations following the procedure described by Korenberg and Chen (1995). BAC 119E16 was confirmed to map on chromosome 21q22.11-12 by reviewing more than 50 cells. This was further confirmed as well by PCR using custom-designed primers for SH3D1A based on sequencing information.

- 10
- 15 **Sequencing cDNA and part of the genomic DNA:** The cDNA was sequenced using RT-PCR products templated on total brain cDNA or directly on BAC 119E16 containing the gene.

- Reverse transcription - polymerase chain reaction (RT-PCT):** SH3D1A cDNA was amplified by RT-PCR using a standard method. Briefly, the control RNA was isolated from a normal male cell line using the TRI reagent kit (Molecular Research Center, Inc. Cincinnati, OH). The first strand of cDNA was then produced using SuperScript Choice System (Pharmacia LKB Biotechnology). The PCR reaction was performed using custom designed primers with PCT-100 Programmable Thermal Controller by a standard PCR procedure. The PCR products for sequencing were prepared by purification with Geneclean Kit (BIO 101, Inc., Vista, CA) prior to sequencing. To produce clearer sequence, some PCR products were subcloned into pCR-2.1 Vector (CLONETECH Laboratory, Inc.) prior to sequencing.
- 20
- 25

- 30 **PCR of genomic DNA:** three genomic (exon) fragments were generated via PCR by using the BAC 119E16 DNA as template, and purified and sequenced as described above

and below.

Sequencing SH3D1A:

The nucleotide sequence of both the coding and non-coding strands were determined in their entirety by the dideoxy chain termination methods using the ABI PRISM Sequences DNA sequencing kit (PERKIN ELMER) with custom-made primers. The template for DNA sequencing were either PCR products or subclones as described above.

Sequencing the upstream region of SH3D1A:

In order to complete sequencing of the 5' end of SH3D1A and identify the site of initiation of transcription, the following two methods were utilized:

1.5' RACE:

5' RACE was performed by using 5' Marathon RACE kit (CLONETECH Laboratories, Inc. CA). The reaction products were then electrophoresed onto 1% of SeaPlaque GTG agarose (FMC BioProducts, Rockland, ME). The products with the longest sizes (>2Kb) were then further confirmed by sequencing nested PCR fragments.

2. cDNA isolation from cDNA library:

The human cDNA clones were obtained from a cDNA library screening as described in Yamakama et al., (1995). The cDNAs were oligo (dT) primed and cloned unidirectionally into the EcoRI and ChoI sites of the vector. The size of the clones were analyzed by electrophoresis and then using for sequencing.

Sequencing Analysis:

Data processing was performed using ABI Sequencing Analysis software which assessed trace quality and assembled sequence data (ABI Autoassemble program). The vector clipping was performed manually. To ensure the accuracy of the sequence, all regions of the finished sequence was covered by more than one subclone or PCR fragments, usually 3-5X and always were sequenced in opposite orientations. The sequence of the human SH3D1A was screened against Genbank (BLASTN & BLASTX). It was also compared with the previously published SH3P17 sequence (Hsu61166) by using V-gcg program. Significant differences between the previously published SH3P17 and this

newly sequenced SH3D1A were found. These equalled about 8% of the nucleotides. Previous sequence totalled only 3,230bps of the 3' end vs. the subject invention's sequence of 5,200bp. Comparison using with the complete homology sequence gb#AF032118 in *Xenopus Leavis* indicated the same protein start site and a similar but not identical domain structure, see Figures 1 and 2.

SH3D1A Gene Structure:

Protein structure was based on cNDA sequence analysis. The four SH3 domains were confirmed previously (Sparks et al., 1996). However, most significant was the definition of additional domains including EH domain (Eps Homolog domain) in the N terminal end that have been associated with protein interactions involved with cell cycle control and morphogenesis. These suggested a possible role, both in human embryogenesis and in cancers, notably the leukemias associated with Down Syndrome (DS), the decreased platelets associated with deletion of chromosome 21 reported by Fannin et al., 1995, and the familial platelet disorder reported by Dowton et al. (1985) and Ho et al. (1996), all of whose map positions include SH3P17.

Gene expression study by Northern Blotting:

Northern blots made from human multiple tissues were used to perform this study according to the manufacturer's instruction (CLONETHch Laboratory, Inc., CA). Referring to Figure 6, the gene was found to be expressed in all adult human tissues tested, those included Heart, brain, placenta, lung, liver, muscle, kidney and pancreas.

Preparation of full length cDNA Clones corresponding to SH3D1A

A cDNA library based on fetal brain was screened in the same manner as described above with respect to the isolation and sequencing of SH3D1A. Accordingly, Sequencing of 5 different sizes of the cDNA clones was conducted, and indicated that there are at least three isoforms that exist. As all of the sequenced cDNA clones shown in Figure 8, #21 was a full-length cDNA that contains 5438 nucleotides and codes for 1221 amino acids; #11 was a shorter full-length cDNA that contains 5179 nucleotides and codes for 1215 amino acids; clone #s 5 and #9 represent 2192bp, 3193bp and 3128bp length cDNA

respectively, while #5 was identical to #21 and #11 at the 5' UTR containing only two EH domains.

The comparison between cDNAs generated in this study vs previously published homologous, or the comparison between each cDNAs isolated in this study, we found significant differences as shown in Figure 18. The differences between #21 vs ITSs, #21 vs #11 and #9 vs SH3P17 are listed here: #21 is 99.8% identical to ITSs (AF064243; Guipponi et al., 1998) at protein level showing only 1 amino acid different at the position of 114, while at the 5' UTR, the extra 160bp and XXbp difference at the 3' UTR of #21 that gives a 96.7% identity at neucleotides level; #11 was missing 5 amino acids at the position of cDNA 2573-2586 within SH3-A domain and missing 222 neucleotides within 3' UTR region while comparing to #21; #9 was 100% identical to SH3P17 (GenBank Hsu61166, Sparks et al., 1996) at coding region, but it shows 76.8% identity at neucleotides level, the major difference is at the 3' UTR, that is a total of 222bp is missing at the position of 2189 (3963-1774) to 2411 and presents at the same position as shown at #11 vs #21. #9 and SH3P17 only showed four SH3 domains missing SH3-C domain (Guipponi et al., 1998) (Figure 3).

The homologies of ITSN to other proteins were also included in Figure 2. (Sparks et al. 1996 and Guipponi et al. 1998) as discussed by Guipponi et al., 1998.

Genomic organization of the ITSN gene and comparison to SH3P17 and ITSs/ITSI:

The comparison of the human SH3D1A to sequenced human genomic DNA (GenBank No AP000050, AP000049 and AP000048) in this region on chromosome 21 revealed that this gene consists of 29 exons (Figure 3 and Table 2 for exact exon-intron boundaries), the sizes of which vary from 44 to 1516 bp. The sizes of the introns range from 355bp to 7.5Kb. All introns have splice donor and acceptor sites that confirm to the general GT-AG consensus motif. The putative SHD1A translation initiation codon is located on exon 2, while the stop codon is on exon 28.

30

Characterization of the 5' upstream sequence

To determine the 5' upstream sequence of the human SH3D1A gene, the sequence from PAC T1276 was used to carry out the analysis for searching the promoter(s).

Complex mRNA expression on multiple adult and fetal tissues (See Figure 17:

5 Summary of studies on ITS)

As shown in the table and figure, Northern blot of SH3D1A on multiple adult and fetal tissues revealed unexpectedly complicated results. A total of 14 probes were used for expression study (Part 1). There were 6 major mRNA transcripts detected, including a 5.4kb of mRNA fragment that was expressed ubiquitously (Heart, brain, placenta, lung, liver, muscle, kidney and pancreas) in adult and fetal tissues (brain, lung, liver and kidney) using any of the probes used as shown in the top portion of the Figure; a 2.5kb fragment expressed in adult ubiquitously, but strong in muscle while using probe #1 (exon 1); a 2.0 kb fragment that was expressed ubiquitously in adult and fetal while using all of the probes except for probes #2, 3 and #12-13 (exon 2-7 and exon 28-29); the strongest expression were shown on muscle in adult and on liver and brain in fetal; a 4.5kb fragment expressed ubiquitously, but stronger on liver, only seen in fetal while using probes #4, 6, 9 and 12 (exon 7 to 17 and exon 23-25; finally, a fragment larger than 11kb that was expressed specifically on brain by using probes #2 and 3 (exons 2 to 7) in adult and fetal tissue, and only seen in adult by using probe #9 (exon 22-28). Further, there was a small fragment 1.0 kb also seen on liver in fetal tissue by using probes #4 and 6 (exon 7 to 17).

RESULTS

The data presented herein confirm the role of the genes of the invention in conditions relating to leukemia as well as neural abnormalities and dysfunctions. As mentioned earlier, the genes are observed as to changes that occur in regions related to leukemia, and in relation to brain abnormalities observed with adult brain. The role of this family of genes in the regulation of both neural and leukemic conditions supports a broad modulatory influence on both development and homeostasis that commends their application in the diagnostic and therapeutic modalities presented herein.

30

This invention may be embodied in other forms or carried out in other ways without

5

Various references have been identified and referred to herein. The disclosures of such references as well as other publications, patent disclosures or documents recited herein, are all incorporated herein by reference in their entireties.